Solar System Small-Body Demographics with the Palomar Transient Factory Survey

Thesis by

Adam Waszczak

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 15th, 2015)

Acknowledgements

Working on the Palomar Transient Factory (PTF) team at Caltech has been an enriching and exciting experience, one that I will always look back upon as a valuable component in my development as a professional.

I first and foremost thank my advisor, Prof. Shri Kulkarni (PI of the PTF project), for having actively reached out to me in early 2010—via phone, months before I even arrived at Caltech—inviting me to come onboard as a new graduate student. Even though I knew little to nothing of planetary science at the time, he managed to convey to me the rich and rewarding science potential of this project, all of which and more proved to be true during my time working for him.

Thanks to Eran Ofek for having taken the time during my first year (and his busy postdoc) to teach me the basics of PTF and MATLAB—none of my research would have been possible without his mentoring and inspiration. I thank him and Oded Aharonson for hosting my visit to Israel in 2012.

Thanks to the scientists and staff at Palomar Observatory, Caltech Optical Observatories & IPAC for supporting PTF—with special thanks to Russ Laher and Frank Masci for their technical support of PTF's asteroid science.

Thanks to Prof. Tom Prince for his appreciable amount of advising, his genuine interest in NEAs and his dedication to PTF.

Thanks to the Keck Institute for Space Studies (KISS) at Caltech for its financial support and for involving me in several of its stimulating workshops. I would also like to acknowledge funding from JPL internal research and development.

Thanks to Rex Chang, Prof. Wing Ip and the others at NCU for hosting my visit to Taiwan in 2014, and for their continued work helping to bolster PTF's reputation

within the field of planetary science.

Thanks to Yi, Mansi and Eric for letting me be lead observer on nearly thirty Palomar 200-inch observing runs! It was a pleasure operating that beautiful telescope and contributing to the PTF team's efforts beyond my own science.

Thanks to Brian Bue, Umaa Rebbapragada, Frank Masci and David Levitan for introducing me to the methods of machine learning, making the career in data science I intend on pursuing possible.

Thanks to the staff of the Planetary Science department and GPS Division for making my time at Caltech productive and my work life enjoyable.

Thanks to my fellow planetary grad students Josh Kammer, Peter Gao, Masha Klescheva, Miki Nakajima, Mike Wong, Henry Ngo, Danielle Piskorz, Patrick Fischer and all the rest for the good times we've had both in the office and out.

Thanks to Mike Line, Alex Lockwood, Konstantin Batygin and Ajay Limaye for showing me just how much fun planetary science, Caltech, and Pasadena in general could be—beginning with my very first visit to the department in 2009 as a prospective student through to the time you graduated. I'll always miss the 'used food' at F&E and long islands at BC you introduced me to.

Thanks to my parents and family for supporting me from the beginning, as I never would have made it to Caltech without you.

Lastly, dearest thanks to my loving fiancée Kelley for enduring the New York to LA long-distance relationship for the past three years while I finished my PhD. I look forward to moving back to the East Coast and our marriage and life together!

Abstract

Observational studies of our solar system's small-body populations (asteroids and comets) offer insight into the history of our planetary system, as these minor planets represent the left-over building blocks from its formation. The Palomar Transient Factory (PTF) survey began in 2009 as the latest wide-field sky-survey program to be conducted on the 1.2-meter Samuel Oschin telescope at Palomar Observatory. Though its main science program has been the discovery of high-energy extragalactic sources (such as supernovae), during its first five years PTF has collected nearly five million observations of over half a million unique solar system small bodies. This thesis begins to analyze this vast data set to address key population-level science topics, including: the detection rates of rare main-belt comets and small near-Earth asteroids, the spin and shape properties of asteroids as inferred from their lightcurves, the applicability of this visible light data to the interpretation of ultraviolet asteroid observations, and a comparison of the physical properties of main-belt and Jovian Trojan asteroids. Future sky-surveys would benefit from application of the analytical techniques presented herein, which include novel modeling methods and unique applications of machine-learning classification. The PTF asteroid small-body data produced in the course of this thesis work should remain a fertile source of solar system science and discovery for years to come.

Contents

Acknowledgements **iii**

Abstract **v**

List of Figures **xi**

List of Tables **xxx**

1 Introduction **1**

2 Main-Belt Comets **7**

2.1 Introduction 7

2.2 Raw transient data 11

2.2.1 Survey overview 11

2.2.2 Candidate-observation quality filtering 12

2.2.3 Sample quality assessment 14

2.3 Known-object extraction 15

2.3.1 Implementation of kd-tree indexing 15

2.3.2 Matching ephemerides against the kd-tree 16

2.3.3 Summary of known small bodies detected 17

2.3.4 Overlap of PTF with the WISE and SDSS data sets 19

2.4 Unknown-object extraction 22

2.4.1 Previous and ongoing PTF small-body discovery work 23

2.4.2 A custom discovery algorithm for main-belt objects 24

2.4.3 Summary of objects discovered 27

2.5 Extended-object analysis: Approach 28
2.5.1 Definition of the extendedness parameter μ 28
2.5.2 Systematic (non-cometary) variation in μ 29
2.5.3 Formalism for interpreting μ 30
2.5.4 A model-μ to describe inert objects 31
2.5.5 Defining a visually-screenable sample 33
2.6 Extended-object analysis: Results 35
2.6.1 A new quasi-Hilda comet: 2011 CR_{42} 36
2.7 Statistical interpretation 41
2.7.1 Bayesian formalism 41
2.7.2 Active MBCs in the *entire* main-belt 42
2.7.3 Active MBCs in the *outer* main-belt 43
2.7.4 Active MBCs in the *low-inclination* outer main-belt 44
2.7.5 Active MBCs among low-i outer main-belt objects observed *near perihelion* ($-45^\circ < \nu < 45^\circ$) 45
2.7.6 Active MBCs in the sub-5 km diameter population 45
2.7.7 Active MBCs among low-albedo (WISE-sampled) objects . . . 46
2.7.8 Active MBCs among C-type (SDSS-sampled) objects 47
2.8 Conclusion 47
2.8.1 Summary 47
2.8.2 Comparison to previous work 49
2.8.3 Future work: Photometric (absolute magnitude) variation as a function of orbital anomaly 50

3 Small Near-Earth Asteroids **51**
3.1 Introduction 51
3.2 Overview of the PTF survey 56
3.2.1 Technical and operational characteristics 56
3.2.2 Previous solar system science with PTF 57
3.2.3 Real-time data reduction at IPAC 57

3.2.4 Detections of known streaking NEAs . . . 58
3.3 Streak-detection process . . . 60
3.3.1 Object detection with `findStreaks` . . . 60
3.3.1.1 Algorithm . . . 60
3.3.1.2 Completeness and contamination . . . 64
3.3.2 Machine-learned classification . . . 68
3.3.2.1 Overview . . . 68
3.3.2.2 Implementation and training . . . 71
3.3.2.3 Post-training performance . . . 74
3.3.3 Web-based screening interface . . . 75
3.4 Follow-up and reporting of discoveries . . . 78
3.4.1 Target-of-opportunity (ToO) requests . . . 78
3.4.2 Initial NEA discoveries . . . 79
3.4.3 Blind real-time recovery of known NEAs . . . 82
3.4.4 Unconfirmed discoveries . . . 82
3.4.5 Artificial satellites . . . 84
3.5 De-biased detection rate . . . 85
3.6 Scaling laws for streaked asteroid detection . . . 87

4 Asteroid Lightcurves **91**
4.1 Introduction . . . 91
4.1.1 Asteroid rotation . . . 91
4.1.2 Asteroid phase functions . . . 95
4.2 Observations . . . 97
4.2.1 Overview of the PTF survey . . . 97
4.2.2 This work's data set . . . 98
4.3 Lightcurve model . . . 101
4.3.1 Rotation component . . . 101
4.3.1.1 Intra-opposition constraint . . . 101
4.3.1.2 Second-order Fourier series . . . 102

4.3.2 Phase-function component . . . 103
4.3.2.1 Two-parameter Shevchenko model . . . 104
4.3.2.2 Lumme-Bowell G model . . . 107
4.3.2.3 Muinonen et al. G_{12} model . . . 108
4.3.2.4 Multi-parameter Hapke model . . . 109
4.4 Lightcurve-fitting algorithm . . . 110
4.4.1 Linear phase-function parameters . . . 112
4.4.2 Nonlinear phase-function parameters . . . 116
4.4.3 Comments on implementation . . . 120
4.5 Reliability of fitted rotation periods . . . 121
4.5.1 Known-period subsample . . . 122
4.5.2 Machine learning . . . 127
4.5.2.1 Classifier training . . . 131
4.5.2.2 Classifier cross-validation . . . 135
4.5.2.3 Machine-vetted lightcurves . . . 138
4.5.3 Manual screening . . . 139
4.5.4 Asteroids with multiple fitted periods . . . 140
4.6 Preliminary lightcurve-based demographics . . . 140
4.6.1 Disclaimer regarding de-biasing . . . 141
4.6.2 Rotation rates and amplitudes . . . 141
4.6.3 Phase-functions and bond albedos . . . 144
4.6.3.1 Taxonomy from lightcurve data . . . 145
4.6.3.2 Wavelength dependence . . . 146
4.6.4 Spins and amplitudes vs. taxonomy . . . 147
4.7 Comparison to MPC-generated magnitudes . . . 149
4.7.1 Filter transformations . . . 149
4.7.2 Absolute magnitudes . . . 151
4.7.3 Predicted apparent magnitudes . . . 151
4.8 Summary . . . 152
4.9 Appendix . . . 155

4.9.1 Multi-survey visible-band color index . . . 155
4.9.2 Compilation of IR-derived diameters . . . 158
4.9.3 Lightcurve data tables . . . 159

5 Asteroids in *GALEX* 174

5.1 Introduction . . . 174
5.2 GALEX asteroid observations . . . 178
5.3 Modeling visible magnitudes . . . 179
5.3.1 (H, G) from MPC data . . . 181
5.3.2 $(D, A_{\rm bond}, G_{12})$ from PTF, infrared, and color data . . . 182
5.3.3 Rotational uncertainty in V . . . 184
5.4 NUV $- V$ color distribution . . . 186
5.5 Albedo vs. wavelength . . . 191
5.6 Comparison to *HST* data . . . 192
5.6.1 Lutetia . . . 192
5.6.2 Ceres . . . 194
5.7 C-type subgroups . . . 195
5.7.1 X complex . . . 198
5.7.2 G types . . . 199
5.7.3 B types . . . 200
5.8 Summary . . . 200

6 Trojan Lightcurve Demographics 201

6.1 Introduction . . . 201
6.2 Color-based taxonomic metric . . . 204
6.3 Rotation periods and amplitudes . . . 206
6.4 Binarity . . . 211
6.5 Photometric phase functions . . . 212
6.6 Summary . . . 215

Bibliography 226

List of Figures

2.1 Distribution of PTF pointings over the first 41 months of operations (March 2009 through July 2012), in sky coordinates relevant to small-body observations. 11

2.2 Transients per image (after filtering and discarding the densest 10% of images) versus ecliptic latitude, off the galactic equator and near-opposition longitude. Vertical gray lines are the scatter (standard deviation) and the black line traces the mean number of detections. Due to the large number of images, the standard error of the mean for each bin is very small (comparable to the width of the black line). The inferred ratio of false positive detections (image artifacts) to real small-body detections at low ecliptic latitudes is at least of order unity. 14

2.3 The shaded gray histograms show the distributions of known objects (normalized such that largest bin equals unity), while the red lines show the fraction of objects in each bin included in the PTF dataset. The osculating orbital elements are from the JPL Small-Body Database, (http://ssd.jpl.nasa.gov), the absolute magnitudes from the Minor Planet Center (http://www.minorplanetcenter.net), the visible albedos from fits to the WISE cryogenic data [Masiero et al. (2011)], and the Sloan colors from the SDSSMOC 4th release [Parker et al. (2008)] supplemented with 2008–2009 data (B. Sesar, personal communication). . . . 17

2.4 Distance residuals of harvested small-body observations with respect to their predicted position. The horizontal axis intentionally extends to 4″, as this is the matching radius we use. Significant contamination due to false-positive detections would increase with distance; this does not appear to be the case. 18

2.5 As in Figure 2.2, the vertical bars show the scatter (standard deviation) and the connected points are the mean values for the bins. Now added in red is the distribution of transients after exclusion of the ∼2 million known-object detections. The original distribution is included, in black, for comparison. 22

2.6 Time interval between consecutive observations of known objects in PTF. This distribution justifies the 48-hour upper limit we impose for tracklet finding, which also was chosen for computational expediency. . 24

2.7 Schematic of our tracklet-finding algorithm, with only one spatial dimension (rather than the actual two) for clarity. Given all transients (gray dots), a kd-tree search rapidly finds those that are nearby in spacetime (within the blue dashed box, of 48-hour full-width) to the central, target transient (green dot). Minimal- and maximal-velocity bounds then define a subset of these (all dots lying in the gray shaded regions, in this case four). All possible past + future pairs are considered (in this case, four possible pairs). Pairs whose predicted midpoint position is sufficiently far from the target transient are immediately rejected (red dashed lines). Pairs with a sufficiently small residual (green dashed lines) then are binned in velocity space, and the velocity bin containing the most pairs is chosen. In this example, however, both non-empty velocity bins have only one pair, in which case the pair with the smallest midpoint residual is chosen. 26

2.8 Of 622 asteroids discovered in multi-night PTF data through July 2012, the MPC has provided orbital solutions for 470 of these. Assuming typical albedos, the smallest ($H \sim 19.5$) objects correspond to ∼0.5-km diameters, while the $H \sim 13$ Trojans correspond to ∼10-km diameters. 28

2.9 Distribution of PTF asteroid observations in the five parameters comprising the observable vector $\vec{o}$. Bottom row: Model-μ values (μ_j-values) plotted as a function of each parameter, while holding all other parameters constant; black bars show the error $\sigma_{\mu,j}$. The red circled point is the value at which the parameter is evaluated in the other plots. These plots only show a small slice of the much larger (and impossible to visualize) five-dimensional gridded function μ_j. 32

2.10 For inclusion in the screening sample, an object's mean extendedness value, $\overline{\delta\mu}$, must exceed zero by more than one-sigma, $\langle\delta\mu\rangle$, which in this histogram is true for all objects to the right of zero. The 111 known (and testable) comets plus the 4 new comets (see Section 2.4.3) are plotted in red; 76 out of these 115 fall to the right of zero, meaning our method is 76/115 = 66% efficient at detecting the known plus new comets comprising our sample. Both of the MBCs in our sample fall to the right of zero, implying that we are 100% efficient at detecting objects at least as extended as these known MBCs. 35

2.11 Like Figure 2.10, but now with each object's median apparent magnitude plotted as well. All small bodies in the sample are represented in the 2D histogram (normalized with respect to each magnitude bin), while the known comets are overplotted as red-white circles, the two MBCs as blue squares. Again, about two-thirds (66%) of the comets lie above zero. This plot suggests that the completeness is expressible as a function of apparent magnitude. That is, $C > 66\%$ for bright comets and $C < 66\%$ for dim comets (approaching zero for > 21 mag), while on average $C = 66\%$. The exact magnitude dependence is sensitive to bin size and is not explored quantitatively here. 37

2.12 Known asteroidally-designated objects whose cometary activity was discovered in PTF in the course of this work. Each image is $2' \times 2'$ (pixel scale $1.01''$). Left: 2011 CR_{42} in g'-band on 2011-03-06. No tail is discernible, but the object's FWHM is twice that of nearby stars. Right: 2010 KG_{43} in R-band on 2010-09-08. A $\sim 1'$-long tail is discernible extending toward the lower left corner of the image. 40

2.13 Known candidate main-belt comets in PTF. Each image is $2' \times 2'$ (pixel scale $1.01''$). Left: P/2010 R2 (La Sagra) in R-band ($R \sim 18.5$ mag) on 2010-08-19, with its tail extending towards the top left. Right: P/2006 VW_{139} in g'-band ($g' \sim 20$ mag) on 2011-12-21, with its two oppositely oriented tails barely discernible by eye. 40

2.14 Probability distributions for estimating the fraction of main-belt comets, based on the results of our sample screening. The $C = 0.66$ case reflects our efficiency with respect to the extendedness distribution of all known comets, while the $C = 1$ case applies to extendedness levels at least as high as the two known candidate MBCs in our sample. In this plot $f_{\min}$ was set to 4×10^{-6}, to facilitate visual comparison of the normalized prior with the normalized posteriors. 43

2.15 Summary of true-anomaly and heliocentric distance coverage of known small bodies in PTF (see Figure 2.3 for other orbital statistics of this sample), from March 2009 through July 2012. Most ($\sim$90%) of objects have only been sampled in at most two 30°-wide true anomaly bins, i.e., less than 1/6 of the orbit. In the left histogram objects are only counted once, but in the rose diagram (middle), each object is counted once for each bin it in which it is sampled (hence the fraction values reported for all twelve bins sum to more than 100%). Although all objects spend more time around aphelion, most only fall above the survey detection limit near perihelion, hence there is a slightly larger fraction of objects observed near perihelion. 48

3.1 Cumulative NEA population distribution models compared to discovered objects. Plot adapted from a figure in [Ruprecht et al. (2014)]. . . 54

3.2 Number of streaked NEA discoveries as a function of time (bins include 1990–2014) and survey, where 'streaked' is here defined as any discovered streak greater than 10 seeing-widths in length (see text for details). . . 54

3.3 Some known small NEAs serendipitously detected by PTF. These observations were retrieved solely by computing these known objects' positions at the epochs of archival PTF images and visually verifying the streak's presence. All images are $200'' \times 200''$ with linear contrast scaling from -0.5σ to 7σ. 55

3.4 Flowchart depicting the PTF streaking-NEA discovery pipeline. 59

3.5 Successful detections (green) and failed detections (red) for both real asteroids and synthetic streaks. Here a 'successful detection' means an object was found by `findStreaks` at the predicted location having a measured length within four streak-widths of the predicted length. In the real data (leftmost plots), multiple detections of unique objects are often very close to one another in the 2D spaces plotted here, such that the total number of points discernible on the plot may appear less than actual. 62

3.6 Distribution of false positive detections from `findStreaks`. The largest concentration of these 'bogus' detections are in the short and faint regime. Structure as a function of orientation angle (bottom) is due to a combination of the correlation sensitivity (see text) and pixel effects, wherein diagonal ($\pm 45°$-oriented) blobs are less likely to exist as their flux is diluted across more pixels. 63

3.7 Examples of streak detections in PTF images. The third column, "assigned pixels", shows the pixels mapped to the object by `findStreaks`, wherein unique objects are distinctly colored. (I) Splitting due to saturated star (undefined pixels on difference image). (J) Splitting due to faintness. (K) Splitting due to bad column in difference image. (L) Extraneous pixels from nearby bright star halo. (M) Missed detection due to near-vertical orientation. (N) Missed detection due to near-horizontal orientation. (O) Missed detection due to large variation in background levels (star halo). (P) Missed detection due to faintness. (Q) Poorly-subtracted star false-positive. (R) Linear radiation hit. (S) Non-linear radiation hit. (T) False positive due to background noise. (U) Isolated segment of longer faint streak (e.g., due to a satellite). (V) Portion of optical ghost artifact. (W) Diffraction spike false positive. (X) Poorly-subtracted galaxy false-positive. 65

3.8 Correlation matrix for the 15 features (descriptions given in Table 3.1) used in the classification process. White squares indicate positive correlation, black indicate negative (anti-) correlation, and the area of each square indicates the magnitude of the correlation. 68

3.9 Importance of each of the 15 features (descriptions given in Table 3.1) used in the classification process. This number represents the fraction of training samples in which each feature contributes more by virtue of being at an earlier node splitting in the decision tree. 69

3.10 Classifier performance for each of the ten cross-validation trials. A plot in true-positive versus false-positive space is commonly referred to as a *receiver operating characteristic* (*ROC*) curve. 69

3.11 In the top row (`findStreaks` only), detection is again defined as the presence of an object whose length is within four streak widths of the true length, as in Section 3.3.2.1. In the bottom row (`findStreaks` plus the classifier), detection is defined as the presence of an object of length within four streak widths of the true length *and* a classification score of $p > 0.4$. 72

3.12 Loss in detection completeness due to the machine classifier—i.e., like Figure 3.10 except considering only those candidates that were first positively detected by `findStreaks`). 72

3.13 Example false positive detections in which the original $200'' \times 200''$ image stamp looks like a real asteroid streak, but the larger field of view clearly indicates the nature of the bogus detection. *Left*: Filament of an optical ghost. *Right*: Glint segment, e.g., from a fast-moving rapidly-rotating piece of space debris. If additional candidates from these larger false-positive objects also appear on the scanning page, their common exposure timestamp implies their stamps will appear adjacent to one another, facilitating their identification as bogus detections. 74

3.14 Discovery images of the first five streaked NEAs found by PTF. 80

3.15 Distribution of PTF exposures (*left*: in sky coordinates, *right*: with respect to opposition) and streaked NEA detections (*right*: with respect to opposition) from 2014-May-01 through 2014-Dec-01. The grayscale scalebar maps the density of PTF exposures in both plots. Exposures for which realtime streak-detection was not performed are not included (e.g., fields lacking reference images or with too high source density on the galactic plane). 80

3.16 PTF streak discoveries that were posted to the NEOCP but never received external follow-up. 83

3.17 Artificial satellites detected as streaks by PTF (identifications provided by the MPC). 84

3.18 Normalized distributions of PTF images and streaked NEA detections with respect to opposition. The 19 NEAs included here consist of new discoveries, blind recoveries and the five unconfirmed discoveries. See Figure 3.15 for the two-dimensional distribution. 85

3.19 Estimates of the number of streaked NEAs detectable by P48 as a function of distance from opposition. Computed using the data in Figure 3.18 and Equation (3.5). 85

4.1 Phase curves (from the literature) containing densely-sampled, rotation-corrected photometry of asteroids in four taxonomic classes. Colored lines are our original fits to the data using various single-parameter ϕ models (cf. Section 4.3.2). 94

4.2 Comparison of predicted asteroid sightings against positive and 'reliable' asteroid detections. We define a 'reliable' detection as any positive detection which (1) lacks any *catalogued* background sources within a $4''$ radius, (2) has a calibrated magnitude uncertainty of less than 0.1 mag, (3) lacks any processing warning flags. As suggested by the middle and right column of plots, this definition of 'reliable' still contains some small contamination (at the <1% level) from uncatalogued background sources and/or noise, as indicated by detections with distance residuals greater than $\sim$1 arcsecond or magnitude residuals of greater than $\sim$1 mag. In panel D, the less than 100% completeness at the bright end reflects the non-negligible probability that any asteroid will fall within $4''$ of a catalogued background source (regardless of the magnitude of either the asteroid or the background source). 96

4.3 Diagram detailing the logic of this work's data reduction and analysis. Includes mining the survey for known-asteroid observations, aggregation of the data into lightcurves, vetting of the lightcurves and an application wherein phase functions are compared to color-derived asteroid taxonomy. See text for details. 114

4.4 Examples of lightcurves having both well-sampled rotation *and* phase-function components. Each row corresponds to a different asteroid. These example asteroids are sorted vertically by their physical diameter (assuming 7% albedo); the top object is ∼45 km and the bottom object is ∼2 km. Column A shows the phase curve (corrected for rotation); Column B shows the rotation curve (corrected for phase-function); Column C shows the periodogram; Column D shows the distribution of the observations in rotational phase vs. solar phase angle. Above each plot is additional information depending on the column: (A) the asteroid number, followed by (in square brackets) the opposition year (most are 2013) and filter (in all cases 'r') followed by the fitted G_{12} parameter; (B) the fitted absolute magnitude and amplitude; (C) the fitted period (in hours); (D) the number of data points included (and shown) in the fit. 119

4.5 For the 927 lightcurves (805 unique asteroids) having a quality code 3 period in the Lightcurve Database of [Warner et al. (2009)] *and* an original fit in this work, we plot the distribution of the relative error in our fitted rotation frequencies with respect to the literature-referenced frequencies. The distribution is bimodal, with the left-hand mode corresponding to those fits having better than ∼3% agreement. 122

4.6 Examples of lightcurves whose fitted frequency differs from the reference frequency by more than 3%, so that they fall in the right mode in the histogram shown in Figure 4.5 and are formally defined as inaccurate fits. *Row 1*: Low-amplitude rotator. *Row 2*: Incorrect period (too few observations?). *Row 3*: A fitted frequency that differs from the reference frequency by 12%. *Row 4*: period that differs by a non-integer multiple, despite looking reasonable. *Row 5*: Folded lightcurve appears to be fitting noise in the data. 123

4.7 *Top row*: The 927-lightcurve known-period sample (black), divided into the accurately-fitted (green) and inaccurately-fitted (red) subgroups. *Middle row*: Ratio of the green to black histograms. *Bottom row*: Results of cross-validation of the machine-classifier (see Section 4.5.2.2). . 124

4.8 *Top row*: The 927-lightcurve known-period sample (black), divided into the accurately-fitted (green) and inaccurately-fitted (red) subgroups. *Middle row*: Ratio of the green to black histograms. *Bottom row*: Results of cross-validation of the machine-classifier (see Section 4.5.2.2). . 125

4.9 Correlation matrices (Spearman's ρ coefficient) for the 20 lightcurve features (Table 4.2) in the training sample (left) and in the full data set (right). 129

4.10 Examples of reliable lightcurves whose folded rotation curve include cusp-like minima (systematic negative deviation from the 2nd-order Fourier fit at minimum brightness), suggestive of a binary system. Many more examples exist in our lightcurves, however in this work we have not specifically flagged such lightcurves. Future works will more carefully label and analyze this particular class of objects. 130

4.11 Definitions of true vs. false and positive vs. negative labels. True-positive rate (TPR) is sometimes called the *completeness* or *sensitivity*, while false-positive rate (FPR) is otherwise known as the *false-alarm rate*, one minus the *reliability*, or one minus the *specificity*. 133

4.12 True-positive versus false-positive rates for the cross-validation trials. Such as plot is sometimes referred to as a *receiver operating characteristic* (ROC) curve. Each trial trains the classifier using a randomly-chosen 80% of the known accurate fits and 80% of the known inaccurate fits among the 927-lightcurves that have reference periods. The 20% remaining lightcurves serve as the test sample. Moving along the hyperbolic locus of points in this plot is equivalent to tuning the classification probability threshold from zero (lower left of the plot) to one (upper right of the plot). The errorbars represent the scatter in the 1,000 cross-validation trials. 134

4.13 Varying the number of features that are randomly split per node in the decision-tree-building process affects both the TPR and FPR. The values plotted here correspond to the $p > 0.5$ classification threshold; each point was generated by the exact same process for which the results in Figure 4.12 were generated, only varying the number of features with respect to which nodes are split. In the left plot, the first four points are labeled with the number of features for that trial (for $n > 4$ we omit the label). In our actual implemented model (Figure 4.12) we chose $n = 4$ features, the value after which the TPR/FPR ratio plateaus at approximately 2, and also the value [Breiman (2001)] recommends, i.e., the square-root of the total number of features (in our case, $\sqrt{20} \approx 4$). 135

4.14 Distributions of PTF-fitted lightcurves (and various subsets thereof) in select features/parameters. These plots are histograms with the same binning as the top rows of Figures 7 and 8. For better readability we here use line-connected bin points (rather than the stair-plot format used in, e.g., Figure 4.5). 162

4.15 Example lightcurves for which the machine-based and human-based reliability scores differ. *Row 1*: Human approved, machine rejected ($p = 0.32$). *Row 2*: Human rejected, machine approved ($p = 0.66$). *Row 3*: Human approved, machine rejected. For this object, the fitted period differs from the known reference period of 392 hours by 7%, hence the machine rejects it by definition. *Row 4*: Human rejected, machine approved ($p = 0.70$). 163

4.16 For the 654 unique asteroids having more than one reliable lightcurve fit (either multiple oppositions and/or both R and g band data) we plot the log of the relative frequency error, defined as the range of the asteroid's fitted periods divided by the geometric mean of its fitted periods. Comparison with Figure 4.5 suggests that we can deem all cases with error $\lesssim$3% as consistently recovered periods, and those with greater than 3% error as inconsistent fits. 164

4.17 *Panel A*: Distribution of spin rate and amplitude as functions of infrared-derived diameters (see appendix for diameter data sources), including data for 4,040 of our lightcurves. The two-dimensional histograms (left side plots) are column-normalized (see text for details). *Panel B*: Comparison of the period versus amplitude distribution (regular 2D histogram, *not* column normalized) with max-spin-rate versus amplitude for a uniform density ellipsoid held together solely by self-gravity . . . 165

4.18 Various fitted phase-function parameters plotted against color index and bold albedo (two-dimensional histograms; the total number of lightcurves in each plot is stated above it as $N = \ldots$). In the right column of plots, one-dimensional distributions with the color-index classified objects plotted separately. In the right column of 1D histograms, C and S types are defined as objects with color indices less than 0.25 and greater than 0.75, respectively. 166

4.19 *Left*: We perform the same clustering analysis used in defining the color index (see appendix), this time on the G_{12} versus $A_{\rm bond}$ distribution, which contains 1,631 PTF lightcurves all of which have IR-derived diameters and reliable phase functions. The output of this clustering analysis is the *photometric index*, which analogous to the color index is a number between 0 (C type) and 1 (S type) quantifying to the class membership of each constituent asteroid data point. *Right*: Correlation between the color index and our photometric index, a comparison which can be made for 361 objects. Note that most data are in the lower left and upper right corners 167

4.20 For the 92 asteroids with both R-band and g-band lightcurve fits from the same opposition, we use the resulting difference in the absolute magnitudes $H_g - H_R$ as a proxy for taxonomy. This color distribution is qualitatively bimodal (top left), and the correlation with G_{12} is very robust (top right). We detect no significant difference in the G_{12} and/or β parameters between the two bands, both in the sample as a whole, and as a function of the $H_g - H_R$ color. 168

4.21 Added completeness from supplementing the color index with the photometric index among asteroids having PTF lightcurves. Both indices are a proxy for the taxonomic type. The left- and right-hand plots apply separately to the subset labeled by the black line above each column. . 169

4.22 Taxonomic dependence on spin rate and amplitude, also versus diameter, using the union of the color-index and photometric-index based C/S taxonomy. 170

4.23 *Left*: Transformations between MPC V band and the PTF R and g bands for asteroids, based on the difference between MPC-fitted and PTF-fitted H magnitudes for asteroids whose PTF-fitted G values are in the range $0.10 < G < 0.20$ as well as other PTF-coverage constraints (see text). *Right*: R-band data only, with S and C types defined with either color and/or photometric indices (again using the <0.25 and >0.75 index thresholds). 170

4.24 Errors in the MPC-listed absolute magnitudes relative to the PTF H values (in R band and using the G_{12} fit's H value), only considering asteroids with IR-derived diameters. On the right is the corresponding geometric albedo relative error. Pixels in the 2D histograms shown here are column normalized. The running-bin geometric mean and 16th and 84th percentiles are shown as green and red lines. Yellow dashed lines are the mean and 84th percentile expected from the 0.1 mag transformation uncertainty alone (for 7% geometric albedo). 171

4.25 Comparison of the root-mean-square residuals, with respect to the PTF (H, G_{12}) plus rotation fit and the MPC (H, G) fit, for all lightcurves having a reliable R-band PTF phase-function fit. 171

4.26 *Panel A*: Fraction of PTF lightcurves with colorimetric data available, for both the reliable-period and reliable-period-plus-G_{12} sets of lightcurves. *Panel B*: Two-dimensional color distributions for seven surveys, and correlations of FCM-clustering-derived classifications between all pairs of surveys. 172

4.27 Relationship between various asteroid surface measurements (from the UV to near-IR) and this work's visible-color-derived C/S color index. See text for descriptions of data the data sets used here, and accompanying references. 173

4.28 Fraction of PTF lightcurves with thermal-IR-based diameter estimates available, for both the reliable-period and reliable-G_{12} sets of lightcurves. 173

5.1 *GALEX* UV and SDSS visible filter response curves (colored lines) convolved with the spectrum of a G2 V type star (black line). The spectrum is from the library of [Pickles (1998)]. The vertical scale is in AB magnitude units per unit wavelength, offset by an arbitrary constant. Note wavelength is plotted on a log scale. 175

5.2 Characteristics of positive asteroid detections from the two *GALEX* surveys (distinguished by exposure time) shown separately in black and green. 178

5.3 We compute $V_{\rm PTF}$ model magnitudes by first assigning fixed $A_{\rm bond}$ and G_{12} values to each *GALEX*-observed asteroid depending on its color class; we then use D to compute H, and finally use the assumed G_{12} value to predict V. The fixed values of $A_{\rm bond}$ and G_{12} are medians from the color-albedo-G_{12} data in [Waszczak et al. (2015)], 2D histograms of which are shown here. Above each plot is the sample size ($N = \ldots$). A total of 642 asteroids have color data *and* G_{12} values in the PTF data; 355 of these also have diameters available (required to compute $A_{\rm bond}$). Panel D shows that *WISE* $W1$ geometric albedos correlate with the PTF bond albedo; we thus use the *WISE* p_{W1} data to assign C types either a low ($A_{\rm bond} \approx 0.01$) or high ($A_{\rm bond} \approx 0.04$) bond albedo. 181

5.4 Distribution of the NUV$-V$ color for *GALEX*-observed asteroids using the H, G model with MPC data to predict V. Plots B and C define C types and S types as objects with color indices of <0.25 and >0.75, respectively. Plots C–F include only the subset with a diameter estimate available; this subset is precisely the same sample considered in Figure 5.6. 185

5.5 Flowchart visualizing the steps in the *GALEX*-observed asteroid sample selection process. Each box is a subset of the box pointing to it. 185

5.6 Distribution of the NUV$-V$ color for *GALEX*-observed asteroids using the D,$A_{\rm bond}$,G_{12} model with PTF data, infrared data, and color data to to predict V. See Table 5.2 for the definitions of the S, $\rm C_{low}$, and $\rm C_{high}$ groups in plots B, C, E, and F. 187

5.7 Various checks for systematic differences in the predicted V magnitudes output by the two different photometric models. *Bottom row*: Investigation of phase-angle-dependence on the NUV $- V$ color. 189

5.8 *Top*: Relative albedo (each type's r-band albedo normalized to unity) versus wavelength for C types and S types. The leftmost (NUV) points are computed from this work's data, the remaining albedos (SDSS bands) are taken directly from [Ivezić et al. (2001)]. *Bottom*: Bandpass response functions (colored lines) convolved with the solar spectrum (black line). 193

5.9 *Top*: *GALEX*, SDSS, and ECAS C-type data from Figure 5.8 compared to *HST*-derived albedos for 1 Ceres and 21 Lutetia. Note the red color here has a different meaning than it does in Figure 5.8. 194

5.10 Visible color/albedo distributions of Tholen-classified [Tholen (1989)] and Bus/Binzel-classified [Bus & Binzel (2002)] C-type subgroups among this work's sample of *GALEX*-observed C-type asteroids. See text for further information. 196

5.11 NUV $- V$ color distributions of Tholen-classified [Tholen (1989)] and Bus/Binzel-classified [Bus & Binzel (2002)] C-type subgroups among this work's sample of *GALEX*-observed C-type asteroids. See text for further information. 197

6.1 *Panel A*: Color distributions for six surveys (diagonal, with blue = C type and orange = D type), and correlations (off diagonal) of FCM-clustering-derived classifications between all pairs of surveys (all pairs of data sets share at least three Trojans). Compare to Figure 5.26 of [Waszczak et al. (2015)]. *Panel B*: Differential number size distribution of Trojans, with subsets relevant to this work. Diameters assume a visible geometric albedo of 0.06 for all Trojans. 217

6.2 A small selection of example Trojan PTF lightcurves. Each row corresponds to a different asteroid. Column A shows the phase curve (corrected for rotation); Column B shows the rotation curve (corrected for phase-function); Column C shows the periodogram; Column D shows the distribution of the observations in rotational phase vs. solar phase angle. Above each plot is additional information depending on the column: (A) the asteroid number, followed by (in square brackets) the opposition year (most are 2013) and filter (in all cases 'r') followed by the fitted G_{12} parameter; (B) the fitted absolute magnitude and amplitude; (C) the fitted period (in hours); (D) the number of data points included (and shown) in the fit. 218

6.3 Distributions of rotation frequency and amplitude for this work's sample of Trojans (combined PTF and literature-referenced data). Diameters assume albedo $p_V = 0.06$; C-type and D-type labels are assigned based on the taxonomic metric (color index) described in Section 5.2. Lines in the bottom row's plots (period-vs.-amplitude) represent the maximum deformation states for fluid bodies of various bulk densities. 219

6.4 Same data as Figure 6.2, except showing only the geocentric mean and 16th–84th percentile (1σ) running statistics as functions of diameter. . 220

6.5 Spin frequency and amplitude distributions (not normalized per datum, as is often done) of large Trojans and main-belt C types. None of the six KS tests noted in this figure represented a statistically significant discrepancy between a pair of distributions. 221

6.6 Distributions of rotation frequency and amplitude for small i.e., $10 < (D/\mathrm{km}) < 40$ size Trojans and main-belt C types. These KS tests do indicate that their observed distributions differ at least at the 3σ significance level. Trojans would probably begin to show effects of YORP in their spin statistics below ∼20 km assuming that main belt C types do around ∼40 km. There could therefore be many sub-20-km Trojans spinning with faster than ∼5 hour periods, simply beyond the diameter limit of this sample. Or, if ∼5 hours is itself the spin limit for Trojans, their observed excess in small slow rotators could also be explained by YORP at these sizes. 222

6.7 PTF lightcurves of six Trojans which were identified as candidate binaries in *WISE* (there were 34 such candidates found by [Sonnett et al. (2015)]). See Figure 6.2 for a description of the plot-labeling format used here. Note the first asteroid (13323) has two lightcurves in PTF, from different oppositions. 223

6.8 PTF lightcurves of two Trojans identified visually, in the course of this work, as likely binary systems. The second object, asteroid 9430, has three different lightcurves: two from the 2012 opposition (R and g band) and one from the 2013 opposition (R band). Note that none of the *WISE* candidates (Figure 6.8) unambiguously show the deep cusp-like minima of the kind seen in the above Trojans. 224

6.9 Phase function parameter distributions of PTF-observed Trojans and main-belt C-type asteroids. Only the 23 Trojans included here had sufficient phase-angle coverage and reliable period (or low amplitude) giving them a 'reliable' phase-parameter estimate. 225

List of Tables

2.1 Known candidate MBCs and impacted asteroids (collectively termed "active main-belt objects") as of April 2013: Summary of orbits and sizes 8

2.2 Known solar system small-body detections in PTF as of July 2012 . . 19

2.3 Known comets observed by PTF. Number of observations and nights; magnitude ranges, heliocentric (r) and geocentric (Δ) distances (in AU). 20

2.3 — *Continued* 21

2.4 Comets discoveries made in PTF in the course of this work 38

2.5 Summary of the PTF comet sample. JF = Jupiter-family; LP = long-period; MB = main-belt. "Observed" means found by the search algorithms of Sections 3 or 4; "model-μ tested" means it lies in the observable ranges shown in Figure 2.9 (e.g., excludes bright comets), and $\overline{\delta\mu} - \langle\delta\mu\rangle > 0$ means positively detected as extended. Objects 2010 KG_{43} and 2011 CR_{42} are counted in all rows as PTF-discovered JFCs, even though they were not included in the $76/115 = 66\%$ efficiency calculation of Section 2.5.5 (since they were not discovered until Section 2.6). 40

3.1 Morphological and other features saved for streak candidates. 62

3.2 PTF discoveries of streaked NEAs (between 2014-May-01 and 2014-Dec-01) 77

3.3 iPTF sub-surveys containing streaked-NEA discovery exposures. 80

3.4 Unconfirmed PTF streak discoveries (from 2014-May-01 to 2014-Dec-01) 83

3.5 Figure of merit comparison for surveys 90

4.1 Description of the PTF asteroid database. Includes PTF data acquired from March 2009 through July 2014, excluding Hα survey data. 99

4.2 Summary of the 20 lightcurve parameters (features) used by our period-quality classifier. See text for a discussion of the cross-validation-derived importance value (Section 4.5.2.2). 128

4.3 Asteroid colorimetry data sets used in computing this work's C/S color metric. These data sets are visualized in Figure 4.26. 154

4.4 Parameters describing PTF lightcurves with a reliable period or phase function. Byte-by-byte Description of file: `ptf_asteroid_lc_parameters.txt` 160

4.5 Parameters describing PTF lightcurves with a reliable period or phase function. Byte-by-byte Description of file: `ptf_asteroid_lc_observations.txt` 161

5.1 Observations of asteroids detected in *GALEX* NUV images. Includes 1,342 detections of 405 asteroids detected at least twice. 176

5.2 $A_{\rm bond}$ and G_{12} (based on PTF data) of color-defined taxonomic groups. 183

5.3 NUV $- V$ color (mag. units) of *GALEX*-observed asteroids and sample sizes 186

6.1 Trojan color data sets used in computing this work's C/D taxonomic metric. These data sets are visualized in Figure 6.1 panel A. 202

Chapter 1

Introduction

The Palomar Transient Factory (PTF) is one of a new generation of synoptic sky-surveys designed to produce 'video footage' of the sky, as compared to previous surveys which only presented single 'snapshots' in time. This is now possible due to the advent of robotic telescopes, efficient CCD imaging technology, and the ability to reliably store and access very large amounts of raw digital data. As the name implies, its primary intended function is to discover and study extragalactic astrophysical *transients*—phenomena such as supernovae, gamma-ray bursts and microlensing events which offer glimpses of extreme conditions and aspects of our universe. Most of the PTF team exploits the survey for this kind of science. A smaller group of individuals use the data to study variable stars, which typically are only bright enough to detect within our own galaxy, but are interesting for their own reasons, such as accurately mapping the structure of the Milky Way (RR Lyrae) or serving as sources of low-frequency gravitational waves (AM cataclysmic variable systems).

Study of solar system objects (asteroids and comets) was not a primary motivation in the design of the PTF survey, though it was recognized at its inception that it would still prove to be a formidable application of its data. The appeal in acquiring serendipitous asteroid observations was noted more than two centuries ago by French astronomer Jean Baptiste Delambre, five years after the discovery of the first asteroid Ceres by Italian astronomer Piazzi:

> We further remark that these four planets [Uranus, Ceres, Pallas, and Juno] were found while searching for something else, and conclude that

> the real way to deserve and to encounter such accidents is to be occupied in some grand undertaking, which in itself is of real use, and keeps us constantly on the route to such discoveries ... for example, to work, as M. Piazzi, to perfect and augment the stellar catalogue, observing each star repeatedly for several days ... to evidence in the long run the planets that could still be confused among the innumerable quantity of very faint stars scattered in the sky.
>
> —*J.B. Delambre, 1806*

While in the early 19th century there was still much appeal for the mere discovery of new minor planets, today the field is primarily concerned with the *characterization* of these objects (with important exceptions being smaller members of the near-Earth asteroid and trans-Neptunian populations). While in-situ observations made by mission-based spacecraft offer exceptional close-up views of asteroids, they have also made clear the fact that minor planets exhibit extreme diversity in their shapes, size, spin properties and mineralogical compositions. Survey-based observations on the other hand facilitate *population-level* science (i.e., demographics) of asteroids, allowing us to understand the extent and causes of this diversity, especially when coupled with modern-day simulations of the formation of the solar system and key events of its history, in which the asteroids act as key test particles.

The Palomar 48-inch Oschin Schmidt Telescope has itself undertaken many minor planet surveys since its construction in the mid 20th century. As some of the main factors determining its sensitivity—its aperture size and optics—have not been altered significantly since then, most of the asteroids observed by it and similar-sized telescopes today have long-since been discovered and their orbits catalogued. The novel *temporal* sampling offered by PTF however allows us to gather sufficient number of observations per asteroid to undertake systematic searches for transient cometary events, or to search for asteroids which are so small that they are only detectable when sufficiently close to the Earth (for a relatively short window of time). It permits the measurement of spin periods and allows us to put constraints on the

elongation of many thousands of asteroids simultaneously—an endeavor which up until recently individual observers would use many allocated nights of telescope time performing targeted observations to achieve.

The content of this thesis is a compilation of five manuscripts, one of which has been published (Chapter 2; [Waszczak et al. (2013b)]), one of which has been accepted for publication and is in press (Chapter 4; [Waszczak et al. (2015)]), and one of which is in review (Chapter 5). The data taken into account in each chapter vary slightly depending upon when the analysis was performed: Chapter 2 considers PTF data collected from 2009–2012 whereas later chapters consider data collected 2009–2014.

In Chapter 2, we search the PTF survey to derive upper limits on the population size of active main-belt comets (MBCs). Cometary activity in main-belt asteroids probes the ice content of these objects and provides clues to the history of volatiles in the inner solar system. From data collected March 2009 through July 2012, we extracted ∼2 million observations of ∼220 thousand known main-belt objects (40% of the known population, down to ∼1-km diameter) and discovered 626 new objects in multi-night linked detections. We formally quantify the "extendedness" of a small-body observation, account for systematic variation in this metric (e.g., due to on-sky motion) and evaluate this method's robustness in identifying cometary activity using observations of 115 comets, including two known candidate MBCs and six newly-discovered non-main-belt comets (two of which were originally designated as asteroids by other surveys). We demonstrate a 66% detection efficiency with respect to the extendedness distribution of the 115 sampled comets, and a 100% detection efficiency with respect to extendedness levels greater than or equal to those we observed in the known candidate MBCs P/2010 R2 (La Sagra) and P/2006 VW_{139}. Using a log-constant prior, we infer 95% confidence upper limits of 33 and 22 active MBCs (per million main-belt asteroids down to ∼1-km diameter), for detection efficiencies of 66% and 100%, respectively.

In Chapter 3, we describe the successful implementation of a real-time system for near-Earth asteroid detection with PTF. Near-Earth asteroids (NEAs) in the 1–100

meter size range are estimated to be ∼1,000 times more numerous than the ∼11,000 currently-catalogued NEAs, most of which are in the 0.5–10 kilometer size range. Impacts from 10–100 meter size NEAs are not statistically life-threatening but may cause significant regional damage, while 1–10 meter size NEAs with low velocities relative to Earth are appealing targets for space missions. We describe the design and initial results of a real-time NEA-discovery system specialized for the detection of small, high angular rate (visually-streaked) NEAs in PTF images. Our real-time NEA discovery pipeline uses a machine-learned classifier to filter a large number of false-positive streak detections, permitting a human scanner to efficiently and remotely identify real asteroid streaks during the night. Upon recognition of a streaked NEA detection (typically within an hour of the discovery exposure), the scanner triggers follow-up with the same telescope and posts the observations to the Minor Planet Center for worldwide confirmation. We describe our ten initial confirmed discoveries, all small NEAs that passed 0.3–15 lunar distances from Earth. Lastly, we derive useful scaling laws for comparing streaked-NEA-detection capabilities of different surveys as a function of their hardware and survey-pattern characteristics. This work most directly informs estimates of the streak-detection capabilities of the Zwicky Transient Facility (ZTF, planned to succeed PTF in 2017), which will apply PTF's current resolution and sensitivity over a 47-deg^2 field-of-view.

In Chapter 4, we fit 54,296 sparsely-sampled asteroid lightcurves in the PTF survey to a combined rotation plus phase-function model. Each lightcurve consists of 20 or more observations acquired in a single opposition. Using 805 asteroids in our sample that have reference periods in the literature, we find the reliability of our fitted periods is a complicated function of the period, amplitude, apparent magnitude and other lightcurve attributes. Using the 805-asteroid ground-truth sample, we train an automated classifier to estimate (along with manual inspection) the validity of the remaining ∼53,000 fitted periods. By this method we find 9,033 of our lightcurves (of ∼8,300 unique asteroids) have 'reliable' periods. Subsequent consideration of asteroids with multiple lightcurve fits indicate a 4% contamination in these 'reliable' periods. For 3,902 lightcurves with sufficient phase-angle coverage and ei-

ther a reliably-fit period or low amplitude, we examine the distribution of several phase-function parameters, none of which are bimodal though all correlate with the bond albedo and with visible-band colors. Comparing the theoretical maximal spin rate of a fluid body with our amplitude versus spin-rate distribution suggests that, if held together only by self-gravity, most asteroids are in general less dense than $\sim$2 g/cm^3, while C types have a lower limit of between 1 and 2 g/cm^3. These results are in agreement with previous density estimates. For 5–20 km diameters, S types rotate faster and have lower amplitudes than C types. If both populations share the same angular momentum, this may indicate the two types' differing ability to deform under rotational stress. Lastly, we compare our absolute magnitudes (and apparent-magnitude residuals) to those of the Minor Planet Center's nominal ($G = 0.15$, rotation-neglecting) model; our phase-function plus Fourier-series fitting reduces asteroid photometric RMS scatter by a factor $\sim$3.

In Chapter 5, we apply results from the previous chapter's modeling to facilitate the interpretation of ultraviolet photometry (NUV band, 180–280 nm) of 405 asteroids observed serendipitously by the *Galaxy Evolution Explorer* (*GALEX*) from 2003–2012. All asteroids in this sample were detected by *GALEX* at least twice. Unambiguous visible-color-based taxonomic labels (C type versus S type) exist for 315 of these asteroids; of these, thermal-infrared-based diameters are available for 245. We derive NUV $- V$ color using two independent models to predict the visual magnitude V at each NUV-detection epoch. Both V models produce NUV $- V$ distributions in which the S types are redder than C types with more than 8σ confidence. This confirms that the S types' redder spectral slopes in the visible remain redder than the C types' into the NUV, this redness being consistent with absorption by silica-containing rocks. The *GALEX* asteroid data confirm earlier results from the *International Ultraviolet Explorer*, which two decades ago produced the only other sizeable set of UV asteroid photometry. The *GALEX*-derived NUV$-V$ data also agree with previously published *Hubble Space Telescope* (*HST*) UV observations of asteroids 21 Lutetia and 1 Ceres. Both the *HST* and *GALEX* data indicate that NUV band is less useful than u band for distinguishing subgroups within the greater population of visible-color-defined C

types (notably, M types and G types).

Finally, in Chapter 6, we again apply the results of Chapter 4 by focusing on a small subset of the PTF lightcurves: the Jovian Trojans. The orbital and physical properties of Jupiter's Trojan asteroids constrain models of their population's origin, as well as our solar system's broader dynamical history. The Trojans' apparent division into two compositionally-distinct taxonomic groups remains unexplained. The redder-colored D-types are spectrally similar to comet nuclei and many outer solar system bodies, while the less-red C-types happen to be one of the two predominant taxonomic types found in the main asteroid belt. In a search for clues regarding the origin and composition of these two groups, how they came to populate Jupiter's Lagrange points and how they have evolved over the last several billions of years, we present an original analysis of the lightcurve-based demographics of the Trojans. The lightcurve sample consists of data from the publicly-available online Lightcurve Database, supplemented with over 100 new Trojan lightcurves from the Palomar Transient Factory survey. We find that, for 40-km sized and larger objects, the Trojan D types and C types are statistically indistinguishable in their spin rate and amplitude distributions; similarly the large Trojans and main-belt C-types are also indistinguishable in these properties. For 10–40-km sized objects, however, the Trojans and main-belt C types do appear to differ significantly, possibly due to unequal contributions from YORP torques and/or cometary outgassing, both of which have been suggested previously by other authors. We present PTF lightcurves of six of the WISE binary candidates, and manually identify two additional PTF lightcurves as likely Trojan binary systems. A subset of the PTF Trojan lightcurves contain reliable fitted phase function parameters, the distributions of which we find to be indistinguishable (between D and C types, and Trojans and main-belt C types).

Chapter 2

Main-Belt Comets

2.1 Introduction

Though often regarded as quiescent rock- and dust-covered small bodies, asteroids can eject material by a variety of physical mechanisms. One subgroup of these *active asteroids* [Jewitt (2012)] are the *main-belt comets* (MBCs), which we define[1] as objects in the dynamically-stable main asteroid belt that exhibit a periodic (e.g., near-perihelion) cometary appearance due to the sublimation of freshly collisionally-excavated ice. Prior to collisional excavation, this ice could persist over the age of the solar system, even in the relatively warm vicinity of ∼3 AU, if buried under a sufficiently thick layer of dry porous regolith ([Schorghofer (2008)]; [Prialnik and Rosenberg (2009)]).

More complete knowledge of the number distribution of ice-rich asteroids as a function of orbital (e.g., semi-major axis) and physical (e.g., diameter) properties could help constrain dynamical models of the early solar system ([Morbidelli et al. (2012)] and refs. therein). Such models trace the evolution of primordially distributed volatiles, including the "snow line" of H_2O and other similarly stratified compounds. Complemented by cosmochemical and geochemical evidence (e.g., [Owen (2008)]; [Albarède (2009)]; [Robert (2011)]), such models explore the possibility of late-stage

[1]Some controversy surrounds the definitions of "main-belt comet", "active asteroid" and "impacted asteroid". While the term *active main-belt object* ([Bauer et al. (2012)]; [Stevenson et al. (2012)]) is the most general, our particular definition and usage of *main-belt comet* is intended to follow that of [Hsieh and Jewitt (2006)], i.e., periodic activity due to sublimating volatiles.

Table 2.1: Known candidate MBCs and impacted asteroids (collectively termed "active main-belt objects") as of April 2013: Summary of orbits and sizes

main-belt comets	asteroid	a (AU)	e	i (deg)	D (km)	perihelia	references
133P/Elst-Pizarro	7968	3.157	0.16	1.4	3.8 ± 0.6	'96,'02,'07	Elst+ '96; Marsden '96; Hsieh+ '04,'09a,'10
176P/LINEAR	118401	3.196	0.19	0.2	4.0 ± 0.4	'05,'11	Hsieh+ '06b,'09a,'11a; Green '06
238P/Read		3.165	0.25	1.3	$\sim$0.8	'05,'10	Read '05; Hsieh+ '09b,'11b
259P/Garradd		2.726	0.34	15.9	0.3 ± 0.02	'08	Garradd+ '08; Jewitt+ '09; MacLennan+ '12
P/2010 R2 (La Sagra)		3.099	0.15	21.4	$\sim$1.4	'10	Nomen+ '10; Moreno+ '11; Hsieh+ '12a
P/2006 VW_{139}	300163	3.052	0.20	3.2	$\sim$3.0	'11	Hsieh+ '12b; Novaković+ '12; Jewitt+ '12
P/2012 T1 (PANSTARRS)		3.047	0.21	11.4	$\sim$2	'12	Hsieh+ '12c
impacted asteroids							
P/2010 A2 (LINEAR)		2.291	0.12	5.3	$\sim$0.12		Birtwhistle+ '10; Jewitt+ '10; Snodgrass+ '10
Scheila	596	2.927	0.17	14.7	113 ± 2		Larson '10; Jewitt+ '11; Bodewitts+ '11
P/2012 F5 Gibbs		3.004	0.04	9.7	< 2.1		Gibbs+ '12; Stevenson+ '12; Moreno+ '12

(post-lunar formation) accretion of Earth's and/or Mars' water from main-belt objects. Some dynamical simulations ([Levison et al. (2009)]; [Walsh et al. (2011)]) suggest that emplacement of outer solar system bodies into the main asteroid belt may have occurred; these hypotheses can also be tested for consistency with a better-characterized MBC population.

For at least the past two decades (e.g., [Luu and Jewitt (1992)]), visible band CCD photometry has been regarded as a viable means of searching for subtle cometary activity in asteroids—spectroscopy being an often-proposed alternative. However, existing visible spectra of MBCs are essentially indistinguishable from those of neighboring asteroids. Even with lengthy integration times, active MBC spectra in the UV and visible lack the bright 388-nm cyanogen (CN) emission line seen in conventional comets [Licandro et al. (2011)]. Near-infrared MBC spectra are compatible with water ice-bearing mixtures of carbon, silicates and tholins but also suffer very low signal-to-noise [Rousselot et al. (2011)]. The larger asteroids Themis (the likely parent body of several MBCs) and Cybele show a 3-μm absorption feature compatible with frost-covered grains, but the mineral goethite could also produce this feature ([Jewitt and Guilbert-Lepoutre (2012)] and refs. therein). The *Herschel Space Observatory* targeted one MBC in search of far-infrared H_2O-line emission, yet only derived an upper limit for gas production [De Val-Borro et al. (2012)]. In general, the low albedo [Bauer et al. (2012)] and small diameter of MBCs ($\sim$km-scale), along with their low activity relative to conventional comets, makes them unfit for spectroscopic discovery and follow-up. Imaging of their sunlight-reflecting dust and time-monitoring of disk-integrated flux, however, are formidable alternatives which motivate the present study.

As of April 2013 there were seven known candidate MBCs (Table 2.1) out of $\sim$560,000 known main-belt asteroids. These seven are regarded as candidates rather than true MBCs because they all lack direct evidence of constituent volatile species, although two (133P and 238P) have shown recurrent activity at successive perihelia. Three other active main-belt objects—P/2010 A2 (LINEAR), 596 Scheila, and P/2012 F5 (Gibbs)—likely resulted from dry collisional events and are thus not con-

sidered to be candidate MBCs. Four of the seven MBCs were discovered serendipitously by individuals or untargeted surveys. The other three were found systematically: the first in the *Hawaii Trails Project* [Hsieh (2009)], in which targeted observations of ∼600 asteroids were visually inspected , and the latter two during the *Pan-STARRS 1* (PS1; [Kaiser et al. (2002)]) survey, by an automated point-spread function analysis subroutine in the PS1 moving-object pipeline [Hsieh et al. (2012b)]. Three of the seven candidate MBCs were originally designated as asteroids, including two of the three systematically discovered ones, which were labeled as asteroids for more than five years following their respective discoveries by the automated NEO surveys LINEAR [Stokes et al. (2000)] and Spacewatch ([Gehrels and Binzel (1984)]; [McMillan (2000)]).

Prior to this work, two additional untargeted MBC searches have been published. [Gilbert and Wiegert (2010)] checked 25,240 moving objects occurring in the *Canada-France-Hawaii Telescope Legacy Survey* [Jones et al. (2006)] using automated PSF comparison against nearby field stars and visual inspection. Their sample, consisting of both known and newly-discovered objects extending down to a limiting diameter of ∼1-km, revealed cometary activity on one new object, whose orbit is likely that of a Jupiter-family comet. [Sonnett et al. (2011)] analyzed 924 asteroids (a mix of known and new, down to ∼0.5-km diameter) observed in the *Thousand Asteroid Light Curve Survey* [Masiero et al. (2009)]. They fit stacked observations to model comae and employed a tail-detection algorithm. While their sample did not reveal any new MBCs, they introduced a solid statistical framework for interpreting MBC searches of this kind, including the proper Bayesian treatment of a null-result.

In this chapter, we first describe the process of extracting observations of known and new solar system small bodies in the PTF survey. We next establish a metric for "extendedness" and a means of correcting for systematic (non-cometary) variation in this metric. We then apply this metric to a screening process wherein individual observations are inspected by eye for cometary appearance. Finally, we apply our results to upper limit estimates of the population size of active main-belt comets.

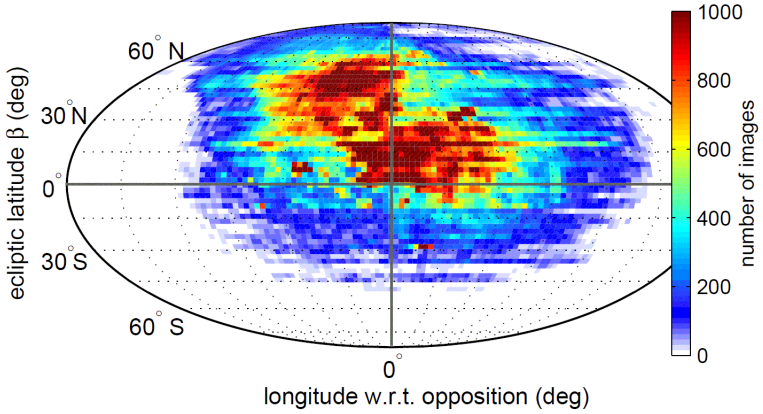


Figure 2.1: Distribution of PTF pointings over the first 41 months of operations (March 2009 through July 2012), in sky coordinates relevant to small-body observations.

2.2 Raw transient data

2.2.1 Survey overview

The *Palomar Transient Factory* (PTF)[2] is a synoptic survey designed to explore the transient and variable sky ([Law et al. (2009)]; [Rau et al. (2009)]). The PTF camera, mounted on Palomar Observatory's 1.2-m $f/2.44$ Oschin Schmidt Telescope, uses 11 CCDs (4096×2048 each) to observe 7.26 deg^2 of the sky at a time with a resolution of $1.01''$/pixel. Most exposures use either a Mould-R or Gunn-g' filter and are 60-s (a small fraction of exposures also comprise an Hα-band survey of the sky). Science operations began in March 2009, with a nominal 2- to 5-day cadence for supernova discovery and typically twice-per-night imaging of fields. Median seeing is $2''$ with a limiting apparent magnitude $R \sim 20.5$ (5σ), while near-zenith pointings under dark conditions routinely achieve $R \sim 21.0$ [Law et al. (2010)].

PTF pointings (Figure 2.1) and cadences are not deliberately selected for solar system science. In fact, PTF's routine sampling of high ecliptic latitudes (to avoid the sometimes bright Moon) alleviates small-body sampling bias with respect to orbital inclination (see Section 2.3.3).

We use data that have been reduced by the PTF photometric pipeline

[2]http://ptf.caltech.edu

([Grillmair et al. (2010)]; [Laher et al. (2014)]) hosted at the *Infrared Processing and Analysis Center* (IPAC) at Caltech. For each image, the pipeline performs debiasing, flat-fielding, astrometric calibration, generation of mask images, and creation of a catalog of point sources using the astrometric reduction software SExtractor [Bertin and Arnouts (1996)]. Code-face parameters such as `MAGERR_AUTO` in this chapter refer to SExtractor output quantities.

Absolute photometric calibration is described in [Ofek et al. (2012a)] and [Ofek et al. (2012b)] and routinely achieves precision of ∼0.02 mag under photometric conditions. In this chapter, we use relative (lightcurve-calibrated) photometry (Levitan et al. in prep; for algorithm details see [Levitan et al. (2011)]), which has systematic errors of 6–8 mmag in the bright (non-Poisson-noise-dominated) regime. Image-level (header) data used in this study were archived and retrieved using an implementation of the Large Survey Database software (LSD, [Jurić (2011)]), whereas detection-level data were retrieved from the PTF photometric database.

2.2.2 Candidate-observation quality filtering

Prior to ingestion into the photometric database, individual sources are matched against a PTF reference image (a deep co-add consisting of at least ∼20 exposures, reaching ≥ 21.7 mag). Any detection not within $1.5''$ of a reference object is classified in the database as a transient. The ensemble of transients forms a raw sample from which we seek to extract asteroid (and potential MBC) observations. As of 2012-Jul-31 there exist ∼30 thousand deep reference images (unique filter-field-chip combinations) against which ∼1.6 million individual epoch images have been matched, producing a total of ∼700 million transients. Of these, we discard transients which satisfy any of the following constraints:

- within $4''$ of a reference object
- outside the convex footprint of the reference image
- from an image with astrometric fit error $> 1''$ relative to the 2MASS survey [Skrutskie et al. (2006)] or systematic relative photometric error > 0.1 mag

- within 6.5 arcmin of a $V < 7$ Tycho-2 star [Høg et al., 2000], the approximate halo radius of very bright stars in PTF
- within 2 arcmin of either a $7 < V < 10$ Tycho-2 star or a $7 < R < 10$ PTF reference source, a lower-order halo radius seen in fainter stars
- within 1 arcmin of a $10 < R < 13$ PTF reference source; most stars in this magnitude range do not have halos but do have saturation and blooming artifacts
- within 30 pixel-columns of a $V < 10$ Tycho-2 star on the same image (targets blooming columns)
- within 30 pixels of the CCD edge
- flagged by the IPAC pipeline as either an aircraft/satellite track, high dark current pixel, noisy/hot pixel, saturated pixel, dead/bad pixel, ghost image, dirt on the optics, CCD-bleed or bright star halo (although the above-described bright star masks are more aggressive than these last two flags)
- flagged by SExtractor as being either photometrically unreliable due to a nearby source, originally blended with another source, saturated, truncated or processed during a memory overflow
- overconcentrated in flux relative to normal-PSF (stellar) objects on the image (i.e. single-pixel radiation hit candidates)—true if the source's `MU_MAX` − `MAG_AUTO` value minus the image's median stellar `MU_MAX` − `MAG_AUTO` value is less than −1 (this criterion is further explained in Section 2.5)

Application of the above filtering criteria reduces the number of transients (moving-object candidates) from ∼700 million to ∼60 million detections. While greatly reduced, this sample size is still too large to search (via the methods outlined in the following sections) given available computing resources—hence we seek to further refine it. These non-small-body detections are likely to include random noise, difficult-to-flag ghost features [Yang et al. (2002)], less-concentrated radiation hits, bright star and galaxy features missed by the masking process, clouds from non-photometric nights, and real astrophysical transients (e.g., supernova).

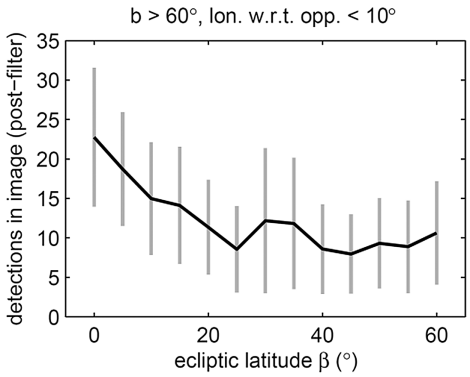


Figure 2.2: Transients per image (after filtering and discarding the densest 10% of images) versus ecliptic latitude, off the galactic equator and near-opposition longitude. Vertical gray lines are the scatter (standard deviation) and the black line traces the mean number of detections. Due to the large number of images, the standard error of the mean for each bin is very small (comparable to the width of the black line). The inferred ratio of false positive detections (image artifacts) to real small-body detections at low ecliptic latitudes is at least of order unity.

We find that about two-thirds of the transients in this sample occur in the densest $\sim$10% of the images (i.e. images with more than $\sim$50 transients). These densest $\sim$10% of images represent over 50% of all images on the galactic equator ($|b| < 20°$), but only 7% of all images on the ecliptic ($|\beta| < 20°$). Hence, discarding them from our sample should not have a significant effect on the number of small-body observations we extract. Discarding these dense images reduces our sample of transients to $\sim$20 million.

2.2.3 Sample quality assessment

Figure 2.2 details the distribution of transients (candidates) per image as a function of ecliptic latitude, after applying the above filters and discarding the dense images. The galactic signal (not shown) is still present: off-ecliptic low galactic latitude fields have a mean of $\sim$40 transients; this number drops roughly linearly with galactic latitude, implying significant residual contribution from ghosts and other missed dense-field artifacts. However, off the galactic equator a factor-of-two increase in the mean number of detections per image is seen from $|\beta| = 50°$ toward the ecliptic, indicating a clear detection of the solar system's main belt.

2.3 Known-object extraction

Having defined our sample of candidate observations, we now seek to match it to objects with known orbits. We first index the candidate observations into a three-dimensional kd-tree, then match this tree against ephemeris data (predicted positions) for all objects. The reader who wishes to skip over the details of the matching algorithm should now go to Section 2.3.3.

2.3.1 Implementation of kd-tree indexing

A kd-tree (short for k-dimensional tree) is a data structure which facilitates efficient cross-matching of M query points against N data points via a multi-dimensional binary search. Whereas a brute force cross-matching involves of order MN computations, a kd-tree reduces this to order $M \log N$. [Kubica et al. (2007)] gives an introduction to kd-trees (including some terminology we use below) and details their increasingly common application in the moving-object processing subsystem (MOPS) of modern sky-surveys.

Our kd-tree has the following features. Since the detections are three dimensional points (two sky coordinates plus one time), the tree's *nodes* are box volumes, each of which is stored in memory as six double precision numbers. Before any leaf nodes (single datum nodes) are reached, the n^{th} level of the tree consists of 2^{n-1} nodes, hence each level of the tree is stored in an array of size $2^{n-1} \times 6$ or smaller.

After definition via median splitting, the bounds of each node are set to those of the smallest volume enclosing all of its data. The splitting of nodes is a parallelized component of the tree-construction algorithm (which is crucial given their exponential increase in number at each successive level). Because the splitting-dimension is cycled continuously, the algorithm will eventually attempt to split data from a single image along the time dimension; when this occurs it simply postpones splitting until the next level (where it is split spatially).

2.3.2 Matching ephemerides against the kd-tree

After constructing the kd-tree of moving-object candidates, we search the tree for known objects. For each of the ∼600 thousand known solar system small bodies we query JPL's online ephemeris generator HORIZONS [Giorgini et al. (1996)] to produce a one-day spaced ephemeris over the 41-month time span of our detections (2009-Mar-01 to 2012-Jul-31). We then search this ephemeris against the kd-tree of candidate detections. In particular, the 1,250 points (days) comprising the ephemeris are themselves organized into a separate (and much smaller) kd-tree-like structure, whose nodes are instead defined by splits exclusively in the time dimension and whose leaf nodes always consist of two ephemeris points spaced one day apart.

The ephemeris tree is "pruned" as it is grown, meaning that at each successive level all ephemeris nodes not intersecting at least one PTF tree node (at the same tree level) are discarded from the tree. Crossing of the R.A. = 0° discontinuity is dealt with by detecting nodes that span nearly 360° in R.A. at sufficiently high tree levels. To account for positional uncertainty in the ephemeris, each ephemeris node is given an 8″ buffer in the spatial dimensions, increasing its volume slightly and ensuring that the ephemeris points themselves never lie exactly on any of the node vertices.

Once the ephemeris tree is grown to only leaf nodes (which are necessarily overlapping some PTF transients), HORIZONS is re-queried for the small-body's position at all unique transient epochs found in each remaining one-day node. Since each leaf node's angular footprint on the sky is of order the square of the object's daily motion (∼10 arcmin^2 for main-belt objects—much less than the size of a PTF image), the number of unique epochs is usually small, on the order of a few to tens. The PTF-epoch-specific ephemerides are then compared directly with the handful of candidate detections in the node, and matches within 4″ are saved as confirmed small-body detections. In addition to the astrometric and photometric data from the PTF pipeline, orbital geometry data from HORIZONS are saved.

Given the candidate sample of ∼20 million transients, for each known object the search takes ∼4 seconds (including the HORIZONS queries, the kd-tree search and

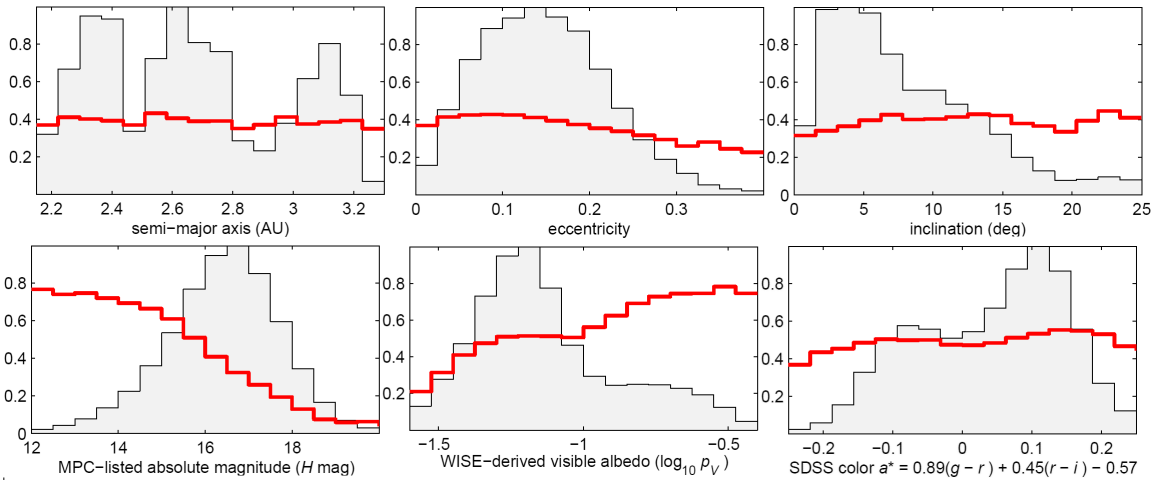


Figure 2.3: The shaded gray histograms show the distributions of known objects (normalized such that largest bin equals unity), while the red lines show the fraction of objects in each bin included in the PTF dataset. The osculating orbital elements are from the JPL Small-Body Database, (http://ssd.jpl.nasa.gov), the absolute magnitudes from the Minor Planet Center (http://www.minorplanetcenter.net), the visible albedos from fits to the WISE cryogenic data [Masiero et al. (2011)], and the Sloan colors from the SDSSMOC 4th release [Parker et al. (2008)] supplemented with 2008–2009 data (B. Sesar, personal communication).

saving of confirmed detections). Hence, PTF observations of the ∼600 thousand known small bodies (main-belt objects, near-Earth objects, trans-Neptunian objects, comets, etc.) require ∼4 days to harvest on an 8-core machine. This relatively quick run time is crucial given that both the list of PTF transients and the list of known small bodies are updated regularly, necessitating periodic re-harvesting.

2.3.3 Summary of known small bodies detected

We used the known small bodies list current as of 2012-Aug-10, consisting of 333,841 numbered objects, 245,696 unnumbered objects, and 3157 comets (including lettered fragments and counting only the most recent-epoch orbital solution for each comet). Our search found 2,013,279 observations of 221,402 known main-belt objects in PTF (∼40% of all known). Table 2.2 details the coverage into various other orbital subpopulations.

Two active known candidate main-belt comets appeared in the sample: P/2010

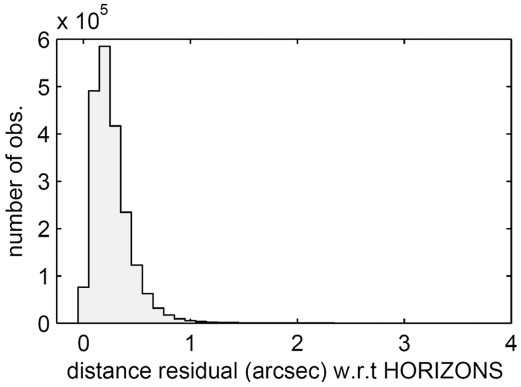


Figure 2.4: Distance residuals of harvested small-body observations with respect to their predicted position. The horizontal axis intentionally extends to 4″, as this is the matching radius we use. Significant contamination due to false-positive detections would increase with distance; this does not appear to be the case.

R2 (La Sagra) was detected 34 times in 21 nights between 2010-Jul-06 and 2010-Oct-29, and P/2006 VW_{139} was detected 5 times in 3 nights—2011-Sep-27, 2011-Oct-02 and 2011-Dec-21 (see Figure 2.13 in Section 2.2.6). In addition to these MBCs, there were 108 known Jupiter-family comets and 65 long-period comets in this sample (see Table 2.3).

In terms of coverage, 54% of the objects are observed five times or fewer, and 53% of the objects are observed on three nights or fewer. Observation-specific statistics on this data set, such as apparent magnitudes, on-sky motions, etc., appear later in Section 2.5 (see the histograms in Figure 2.9). Lastly, a summary of orbital-coverage statistics appears in Figure 2.15 (Section 2.7).

The orbital distribution of the PTF sample is shown in the top row of Figure 2.3. The fraction of known objects sampled appears very nearly constant at 40% across the full main-belt ranges of the orbital elements a, e and i. With respect to absolute magnitude (referenced for all objects from the Minor Planet Center), the PTF sampling fraction of 40% applies to the $H \sim 17$ mag bin, corresponding to 1-km diameter objects for a typical albedo of ∼10%.

As shown in Figure 2.4, the distribution of astrometric residuals with respect to the ephemeris prediction is sharply concentrated well within the matching threshold of 4″. Were this data set significantly contaminated by randomly distributed false-

Table 2.2: Known solar system small-body detections in PTF as of July 2012

	main-belt	Trojan & Hilda	comets*	NEOs	TNOs & centaurs
detections	2,013,279	50,056	2,181	6,586	790
objects	221,402	5,259	175†	1,257	75
% of known	39%	55%	3%	13%	4%

*See Table 2.5 for a more detailed breakdown by comet dynamical type.
†The count of 175 comets given here differs from the count of 115 given in the abstract, for various reasons described in Section 2.5.4 and Table 2.5.

positives, then their number would increase with matching distance (i.e., with annular area per unit matching radius), which evidently is not the case.

2.3.4 Overlap of PTF with the WISE and SDSS data sets

During its full-cryogenic mission in 2010, the *Wide-field Infrared Survey Explorer* (WISE; [Wright et al. (2010)]; [Masiero et al. (2011)], [Masiero et al. (2012)]; [Mainzer et al. (2012)]), observed 94,653 asteroids whose model-derived visible albedos (p_V) have errors of less than 0.05, and nearly half (45,321) of these were also observed by PTF. Relative to this known-albedo sample, PTF detected 47% of the dark ($p_V < 0.1$) and ∼69% of the bright ($p_V > 0.1$) asteroids (Figure 2.3, middle bottom).

Of the asteroids that were observed by the *Sloan Digital sky-survey* (SDSS; [York et al. (2000)]) during its 1998–2009 imaging phase, 142,774 known objects have g, r and i photometry with errors of less than 0.2 mag in all three bands, and more than half (72,556) of these objects were also observed by PTF. These data come from the SDSS Moving Object Catalog 4th release [Parker et al. (2008)], which includes data through March 2007, supplemented with more recent SDSS moving object data from 2008–2009 (B. Sesar, personal communication). The principal component color $a^* = 0.89(g-r)+0.45(r-i)-0.57$ is useful for broad (C-type vs. S-type) taxonomic classification ([Ivezić et al. (2002)]; [Parker et al. (2008)]). Relative to this known-color sample, PTF detected 49% of the carbonaceous-colored ($a^* < 0$) and 52% of the stony-colored ($a^* > 0$) asteroids (Figure 2.3, bottom right).

Table 2.3: Known comets observed by PTF. Number of observations and nights; magnitude ranges, heliocentric (r) and geocentric (Δ) distances (in AU).

name	obs.	nights	first date	last date	$V_{\rm min}$	$V_{\rm max}$	$r_{\rm min}$	$r_{\rm max}$	$\Delta_{\rm min}$	$\Delta_{\rm max}$
7P/Pons-Winnecke	5	3	2009-09-15	2009-09-21	20.3	21.0	3.5	3.5	2.9	2.9
9P/Tempel 1	1	1	2010-03-16	2010-03-16	20.0	20.0	2.9	2.9	2.3	2.3
19P/Borrelly	3	3	2009-05-14	2009-06-25	15.6	17.4	3.1	3.4	2.6	3.3
29P/Schwassmann-Wachmann 1	15	13	2011-01-10	2011-02-14	14.4	15.5	6.2	6.2	5.3	5.7
30P/Reinmuth 1	1	1	2010-02-26	2010-02-26	15.3	15.3	1.9	1.9	1.5	1.5
31P/Schwassmann-Wachmann 2	26	15	2009-12-17	2011-04-11	18.1	19.0	3.5	3.6	2.5	2.8
33P/Daniel	1	1	2009-03-18	2009-03-18	16.8	16.8	2.8	2.8	2.0	2.0
36P/Whipple	45	14	2011-09-05	2011-11-22	17.9	19.4	3.1	3.1	2.1	2.5
47P/Ashbrook-Jackson	4	3	2009-12-03	2009-12-11	17.3	17.6	3.3	3.3	2.3	2.4
48P/Johnson	5	3	2010-04-11	2011-06-11	16.2	21.4	2.4	3.8	1.6	2.8
49P/Arend-Rigaux	30	12	2012-01-20	2012-07-16	16.0	20.0	1.7	2.9	1.0	3.1
54P/de Vico-Swift-NEAT	3	2	2009-07-23	2009-07-29	20.1	20.7	2.4	2.4	1.4	1.4
59P/Kearns-Kwee	1	1	2010-02-17	2010-02-17	20.5	20.5	3.4	3.4	2.5	2.5
64P/Swift-Gehrels	3	3	2009-10-03	2010-02-14	15.8	18.6	1.9	2.9	1.9	2.2
65P/Gunn	9	4	2009-05-25	2012-02-05	14.0	19.7	2.9	4.1	2.5	4.4
71P/Clark	8	5	2011-01-12	2011-01-27	19.6	20.6	3.0	3.1	2.2	2.4
74P/Smirnova-Chernykh	6	3	2010-02-16	2010-02-23	16.1	16.5	3.6	3.6	3.0	3.1
77P/Longmore	74	4	2009-03-13	2009-04-06	14.2	14.7	2.4	2.4	1.4	1.5
78P/Gehrels 2	11	3	2011-11-02	2011-11-09	12.3	12.5	2.1	2.1	1.3	1.3
94P/Russell 4	5	4	2010-02-16	2010-06-03	16.2	17.5	2.3	2.3	1.3	2.1
103P/Hartley 2	18	11	2010-06-06	2010-08-03	14.6	18.3	1.5	2.1	0.7	1.5
10P/Tempel 2	6	5	2009-06-09	2010-08-10	10.6	20.0	1.5	3.4	0.7	3.3
116P/Wild 4	5	3	2009-03-27	2009-04-01	13.6	14.0	2.3	2.3	1.6	1.6
117P/Helin-Roman-Alu 1	29	13	2010-02-13	2011-12-04	18.8	19.8	4.5	5.1	4.5	4.8
118P/Shoemaker-Levy 4	3	2	2010-02-18	2010-04-08	13.6	14.7	2.0	2.1	1.3	1.9
123P/West-Hartley	1	1	2010-09-13	2010-09-13	20.3	20.3	3.0	3.0	2.9	2.9
127P/Holt-Olmstead	5	3	2009-08-13	2009-11-17	17.1	18.7	2.2	2.2	1.3	1.6
130P/McNaught-Hughes	2	2	2010-04-11	2010-04-16	20.8	21.0	3.5	3.5	2.5	2.5
131P/Mueller 2	8	2	2011-11-02	2011-11-03	18.6	19.0	2.5	2.5	1.6	1.6
142P/Ge-Wang	3	2	2010-10-03	2010-10-17	20.6	20.6	2.7	2.7	1.7	1.8
143P/Kowal-Mrkos	2	1	2010-08-24	2010-08-24	19.8	20.1	3.7	3.7	2.9	2.9
149P/Mueller 4	17	10	2010-02-16	2010-06-03	18.7	20.1	2.7	2.7	1.8	2.2
14P/Wolf	1	1	2009-11-03	2009-11-03	19.4	19.4	3.1	3.1	2.2	2.2
157P/Tritton	2	2	2009-09-10	2009-11-07	17.2	18.5	1.8	2.2	1.0	1.2
158P/Kowal-LINEAR	17	7	2012-07-22	2012-07-29	18.8	19.4	4.6	4.6	4.1	4.2
160P/LINEAR	7	2	2010-03-28	2012-07-18	18.7	19.2	2.1	5.3	1.4	4.3
162P/Siding Spring	16	10	2010-11-13	2012-03-21	18.8	20.5	2.7	4.7	3.0	3.8
163P/NEAT	3	1	2011-11-03	2011-11-03	19.9	20.2	2.4	2.4	1.5	1.5
164P/Christensen	4	3	2011-09-04	2011-09-08	17.9	18.8	1.9	1.9	2.6	2.6
167P/CINEOS	17	15	2009-06-24	2010-10-29	20.7	21.6	13.9	14.5	13.0	14.1
169P/NEAT	1	1	2009-07-07	2009-07-07	19.0	19.0	2.2	2.2	1.3	1.3
188P/LINEAR-Mueller	1	1	2010-02-19	2010-02-19	21.2	21.2	4.9	4.9	4.0	4.0
202P/Scotti	3	1	2009-03-17	2009-03-17	19.5	19.8	2.5	2.5	2.5	2.5
203P/Korlevic	3	2	2011-01-01	2011-01-08	17.6	18.4	3.6	3.6	2.9	3.0
213P/Van Ness	2	2	2012-01-04	2012-01-05	17.2	17.3	2.5	2.5	2.6	2.6
215P/NEAT	14	6	2011-11-02	2012-01-21	18.1	19.7	3.8	3.9	2.9	4.0
217P/LINEAR	18	10	2009-06-26	2010-03-28	10.4	18.8	1.2	2.5	0.6	2.4
218P/LINEAR	5	2	2009-05-25	2009-05-27	19.1	19.7	1.7	1.7	0.9	0.9
219P/LINEAR	40	12	2010-08-13	2010-11-08	17.4	19.2	2.6	2.8	1.8	2.4
220P/McNaught	3	1	2009-06-01	2009-06-01	19.9	20.5	2.3	2.3	1.4	1.4
221P/LINEAR	7	6	2009-08-13	2009-09-20	20.4	21.2	2.5	2.6	1.6	1.7
223P/Skiff	8	5	2010-08-13	2010-09-03	19.4	20.3	2.4	2.4	1.9	2.1
224P/LINEAR-NEAT	1	1	2009-09-14	2009-09-14	21.4	21.4	2.3	2.3	1.3	1.3
225P/LINEAR	2	2	2009-10-22	2009-10-22	20.2	20.6	1.5	1.5	0.9	0.9
226P/Pigott-LINEAR-Kowalski	2	2	2009-10-16	2009-10-16	19.3	19.6	2.3	2.3	2.2	2.2
228P/LINEAR	12	10	2010-12-31	2012-03-05	18.0	20.4	3.5	3.5	2.6	3.3
229P/Gibbs	4	2	2009-08-19	2009-08-23	19.7	20.2	2.4	2.4	2.1	2.1
22P/Kopff	7	4	2009-06-26	2009-08-02	11.9	12.3	1.6	1.7	0.8	0.9
230P/LINEAR	5	3	2009-12-03	2010-01-12	18.4	18.8	1.9	2.1	1.4	1.6
234P/LINEAR	1	1	2009-12-15	2009-12-15	20.6	20.6	2.9	2.9	3.1	3.1
236P/LINEAR	19	14	2010-06-17	2011-01-25	17.1	20.7	1.9	2.2	0.9	1.9
237P/LINEAR	27	15	2010-07-05	2010-10-02	19.6	21.2	2.8	3.0	2.0	2.3
240P/NEAT	7	5	2010-07-25	2010-12-08	14.5	16.6	2.1	2.2	1.3	2.6
241P/LINEAR	8	8	2010-12-28	2011-02-01	17.4	18.4	2.4	2.6	1.6	1.7
242P/Spahr	16	13	2010-08-15	2011-08-28	19.3	21.1	4.1	4.8	3.7	4.4
243P/NEAT	5	3	2011-11-03	2011-11-22	20.2	20.9	2.9	3.0	2.1	2.2
244P/Scotti	29	17	2010-09-10	2011-01-06	19.3	20.3	4.2	4.3	3.3	3.9
245P/WISE	9	6	2010-07-25	2010-09-11	19.1	20.4	2.5	2.7	1.7	1.7
246P/NEAT	14	7	2011-02-13	2011-11-23	16.5	19.1	3.6	4.3	3.4	4.1
247P/LINEAR	12	7	2010-10-08	2010-11-12	17.1	20.2	1.6	1.8	0.7	1.1
248P/Gibbs	25	15	2010-09-18	2010-12-11	18.2	19.8	2.2	2.5	1.4	1.7
250P/Larson	3	2	2010-11-03	2010-11-04	20.4	20.8	2.2	2.2	2.1	2.2
253P/PANSTARRS	4	3	2011-11-02	2011-12-08	16.9	18.0	2.0	2.0	1.3	1.6
254P/McNaught	1	1	2011-11-03	2011-11-03	17.8	17.8	3.7	3.7	3.0	3.0
260P/McNaught	9	3	2012-07-27	2012-07-30	14.2	14.7	1.6	1.6	0.9	0.9
261P/Larson	18	9	2012-06-25	2012-07-06	19.2	20.3	2.3	2.3	1.7	1.8
279P/La Sagra	6	3	2009-07-20	2009-08-02	20.1	21.4	2.2	2.2	1.3	1.4
P/2006 VW_{139}	5	3	2011-09-27	2011-12-21	19.0	19.6	2.5	2.6	1.5	1.9
P/2009 O3 (Hill)	7	5	2009-09-20	2009-11-07	17.5	18.5	2.7	2.9	1.8	2.0
P/2009 Q1 (Hill)	5	3	2009-08-01	2010-12-31	18.4	19.9	2.8	4.5	2.0	3.7
P/2009 Q4 (Boattini)	6	4	2009-12-16	2010-03-16	13.4	17.8	1.4	1.8	0.6	0.9
P/2009 Q5 (McNaught)	2	1	2009-08-21	2009-08-21	17.0	17.1	2.9	2.9	2.2	2.2
P/2009 SK_{280} (Spacewatch-Hill)	5	3	2009-10-23	2009-11-09	19.8	20.4	4.2	4.2	3.2	3.3

Table 2.3: — *Continued*

name	obs.	nights	first date	last date	$V_{\min}$	$V_{\max}$	$r_{\min}$	$r_{\max}$	$\Delta_{\min}$	$\Delta_{\max}$
P/2009 T2 (La Sagra)	8	6	2009-08-24	2010-03-13	16.5	20.5	1.8	2.3	1.1	1.9
P/2009 WX_{51} (Catalina)	6	1	2009-12-17	2009-12-17	17.4	18.0	1.1	1.1	0.2	0.2
P/2010 A3 (Hill)	3	2	2009-09-13	2010-03-25	16.1	21.3	1.6	2.7	1.8	1.9
P/2010 A5 (LINEAR)	4	3	2010-01-12	2010-02-24	16.3	17.4	1.8	2.0	1.2	1.8
P/2010 B2 (WISE)	1	1	2010-02-23	2010-02-23	20.0	20.0	1.7	1.7	1.1	1.1
P/2010 D2 (WISE)	1	1	2010-03-17	2010-03-17	19.9	19.9	3.7	3.7	3.6	3.6
P/2010 E2 (Jarnac)	5	4	2010-06-08	2010-06-28	19.1	20.2	2.5	2.5	2.0	2.3
P/2010 H2 (Vales)	11	6	2010-04-16	2010-06-01	11.8	15.1	3.1	3.1	2.1	2.4
P/2010 H5 (Scotti)	18	10	2010-05-30	2010-06-17	20.4	21.2	6.0	6.0	5.4	5.6
P/2010 N1 (WISE)	2	2	2010-03-12	2010-03-15	20.9	21.1	2.1	2.1	1.2	1.2
P/2010 P4 (WISE)	3	3	2010-09-15	2010-10-04	20.1	20.8	2.0	2.0	1.2	1.3
P/2010 R2 (La Sagra)	34	21	2010-07-06	2010-10-29	18.2	20.1	2.6	2.7	1.7	2.1
P/2010 T2 (PANSTARRS)	7	5	2010-09-05	2010-09-15	19.9	21.4	4.0	4.0	3.1	3.2
P/2010 TO_{20} (LINEAR-Grauer)	1	1	2010-08-24	2010-08-24	18.7	18.7	5.3	5.3	4.4	4.4
P/2010 U1 (Boattini)	62	33	2009-06-25	2010-11-13	19.1	21.4	4.9	5.0	4.0	4.9
P/2010 U2 (Hill)	34	19	2010-09-07	2010-12-13	17.6	19.9	2.6	2.6	1.6	1.9
P/2010 UH_{55} (Spacewatch)	24	10	2010-10-15	2011-10-17	18.7	20.3	3.0	3.2	2.1	3.6
P/2010 WK (LINEAR)	9	5	2010-08-14	2010-09-22	18.7	20.4	1.8	1.9	1.2	1.6
P/2011 C2 (Gibbs)	28	19	2010-12-02	2012-02-01	19.7	21.1	5.4	5.6	4.7	5.3
P/2011 JB_{15} (Spacewatch-Boattini)	2	2	2010-06-06	2010-06-09	20.7	21.1	5.6	5.6	5.0	5.0
P/2011 NO_1 (Elenin)	1	1	2011-07-30	2011-07-30	19.7	19.7	2.6	2.6	1.6	1.6
P/2011 P1 (McNaught)	4	2	2011-09-08	2011-09-20	18.7	19.4	5.3	5.3	4.6	4.7
P/2011 Q3 (McNaught)	38	20	2011-07-23	2011-11-30	18.5	20.5	2.4	2.5	1.4	2.0
P/2011 R3 (Novichonok)	5	3	2011-10-08	2011-10-10	18.3	18.6	3.7	3.7	2.7	2.7
P/2011 U1 (PANSTARRS)	18	7	2011-11-24	2012-01-18	19.1	20.9	2.6	2.8	1.8	1.9
P/2011 VJ_5 (Lemmon)	5	3	2012-02-04	2012-03-25	18.5	19.4	1.6	1.9	0.9	0.9
P/2012 B1 (PANSTARRS)	9	6	2011-12-11	2012-01-04	19.2	19.9	4.9	4.9	4.0	4.3
C/2002 VQ_{94} (LINEAR)	3	2	2009-06-09	2009-07-06	18.6	18.7	10.2	10.3	9.5	10.0
C/2005 EL_{173} (LONEOS)	2	1	2009-07-28	2009-07-28	19.5	20.4	8.0	8.0	7.3	7.3
C/2005 L3 (McNaught)	47	19	2009-05-12	2012-03-16	14.0	19.4	6.6	11.6	6.0	10.8
C/2006 OF_2 (Broughton)	89	5	2010-01-25	2010-02-16	16.0	16.6	5.5	5.7	4.6	4.7
C/2006 Q1 (McNaught)	11	6	2009-05-13	2009-08-16	14.0	15.1	4.2	4.8	3.5	4.8
C/2006 S3 (LONEOS)	37	25	2009-06-25	2010-09-12	15.4	17.5	6.7	8.9	5.8	8.6
C/2006 U6 (Spacewatch)	4	2	2009-03-25	2010-03-16	16.2	19.7	3.9	6.6	3.0	5.7
C/2007 D1 (LINEAR)	6	3	2010-03-12	2011-03-16	17.8	18.8	10.5	11.7	9.6	10.9
C/2007 G1 (LINEAR)	7	5	2010-12-28	2011-01-23	19.0	20.1	7.5	7.7	6.7	7.1
C/2007 M1 (McNaught)	14	8	2009-07-04	2010-03-17	18.6	20.5	7.7	8.3	7.1	8.0
C/2007 N3 (Lulin)	2	1	2009-12-27	2009-12-27	16.3	16.5	4.6	4.6	3.6	3.6
C/2007 Q3 (Siding Spring)	33	20	2009-11-03	2010-07-23	11.1	14.6	2.3	3.8	2.2	3.9
C/2007 T5 (Gibbs)	7	5	2009-05-16	2009-06-29	19.9	20.5	5.0	5.2	4.5	5.1
C/2007 U1 (LINEAR)	21	13	2009-06-24	2009-09-07	16.7	17.8	4.4	4.9	3.9	4.4
C/2007 VO_{53} (Spacewatch)	26	11	2011-06-24	2012-06-27	17.8	20.7	5.8	7.6	5.5	7.0
C/2008 FK_{75} (Lemmon-Siding Spring)	19	11	2009-06-28	2010-09-28	15.3	16.7	4.5	5.8	4.1	5.2
C/2008 N1 (Holmes)	13	7	2009-07-21	2010-06-06	16.3	18.3	2.8	3.8	2.6	3.6
C/2008 P1 (Garradd)	4	2	2009-08-23	2009-09-14	15.5	15.8	3.9	3.9	3.0	3.2
C/2008 Q1 (Maticic)	15	7	2009-05-13	2011-02-22	16.1	18.8	3.2	7.5	2.6	6.7
C/2008 Q3 (Garradd)	2	1	2010-03-26	2010-03-26	17.6	17.8	3.7	3.7	3.1	3.1
C/2008 S3 (Boattini)	33	12	2010-09-29	2010-11-06	17.4	18.4	8.1	8.2	7.2	7.3
C/2009 F1 (Larson)	2	1	2009-03-27	2009-03-27	18.7	18.8	2.1	2.1	1.2	1.2
C/2009 F2 (McNaught)	2	2	2012-06-26	2012-06-28	20.8	20.8	8.7	8.7	8.1	8.1
C/2009 K2 (Catalina)	19	10	2009-05-08	2009-08-24	17.8	19.8	3.6	4.1	3.6	3.9
C/2009 K5 (McNaught)	3	3	2010-09-27	2010-11-12	13.7	14.4	2.5	3.0	2.3	2.6
C/2009 O2 (Catalina)	4	2	2009-06-29	2009-07-21	19.8	21.3	3.7	4.0	2.7	3.2
C/2009 P1 (Garradd)	7	4	2011-07-21	2012-02-02	8.6	9.5	1.6	2.6	1.4	1.7
C/2009 P2 (Boattini)	44	26	2009-08-13	2010-09-14	18.5	19.8	6.6	6.7	5.7	6.8
C/2009 T3 (LINEAR)	1	1	2010-06-03	2010-06-03	18.9	18.9	2.8	2.8	2.9	2.9
C/2009 U3 (Hill)	20	4	2010-01-17	2010-05-04	16.2	16.7	1.5	1.7	1.3	1.4
C/2009 U5 (Grauer)	5	4	2010-12-08	2011-01-12	20.2	21.1	6.2	6.3	5.7	6.1
C/2009 UG_{89} (Lemmon)	61	32	2011-04-27	2012-04-29	17.0	20.2	4.1	5.7	3.6	5.2
C/2009 Y1 (Catalina)	14	8	2009-12-30	2011-09-28	15.2	19.4	3.5	4.7	2.7	4.2
C/2010 B1 (Cardinal)	5	3	2010-01-11	2010-01-25	17.8	18.0	4.7	4.7	4.0	4.1
C/2010 D4 (WISE)	32	21	2009-05-18	2010-09-18	19.8	21.3	7.2	7.8	6.5	8.2
C/2010 DG_{56} (WISE)	13	9	2010-07-26	2010-09-11	18.1	20.0	1.9	2.2	1.1	1.6
C/2010 E5 (Scotti)	3	2	2010-03-19	2010-03-25	19.8	19.9	4.0	4.0	3.0	3.0
C/2010 F1 (Boattini)	11	9	2009-11-09	2010-01-17	18.5	19.5	3.6	3.6	3.0	3.7
C/2010 G2 (Hill)	10	7	2010-06-23	2012-01-15	12.4	18.9	2.5	5.0	2.1	4.5
C/2010 G3 (WISE)	37	24	2009-10-03	2011-06-26	18.6	20.4	4.9	5.9	4.7	6.3
C/2010 J1 (Boattini)	4	2	2010-06-13	2010-06-24	18.1	19.2	2.3	2.4	1.7	2.0
C/2010 J2 (McNaught)	6	4	2010-06-27	2011-06-10	16.9	20.0	3.4	4.8	2.6	4.2
C/2010 L3 (Catalina)	53	34	2009-08-03	2012-07-16	18.8	20.9	9.9	10.4	9.3	10.3
C/2010 R1 (LINEAR)	27	11	2012-06-01	2012-06-27	16.9	17.4	5.6	5.6	4.8	5.1
C/2010 S1 (LINEAR)	1	1	2010-02-18	2010-02-18	20.3	20.3	9.9	9.9	9.8	9.8
C/2010 U3 (Boattini)	6	5	2010-10-17	2011-09-04	20.0	20.6	17.1	18.4	16.6	17.5
C/2010 X1 (Elenin)	23	18	2011-01-06	2011-02-22	17.6	19.4	3.3	3.9	2.4	3.6
C/2011 A3 (Gibbs)	22	9	2011-03-04	2011-04-15	16.6	17.5	3.5	3.8	2.7	3.1
C/2011 C1 (McNaught)	1	1	2011-08-25	2011-08-25	19.7	19.7	2.3	2.3	1.7	1.7
C/2011 C3 (Gibbs)	1	1	2011-02-11	2011-02-11	20.4	20.4	1.7	1.7	1.3	1.3
C/2011 F1 (LINEAR)	27	18	2010-10-12	2012-05-24	13.4	19.7	3.3	8.2	2.9	8.6
C/2011 G1 (McNaught)	5	2	2011-11-08	2012-01-27	17.3	17.5	2.2	2.6	1.8	2.5
C/2011 J3 (LINEAR)	1	1	2011-06-23	2011-06-23	19.0	19.0	2.4	2.4	2.1	2.1
C/2011 L3 (McNaught)	18	10	2011-07-15	2011-10-01	14.9	17.6	1.9	2.0	1.0	1.9
C/2011 M1 (LINEAR)	2	1	2011-07-04	2011-07-04	15.7	16.0	1.4	1.4	1.4	1.4
C/2011 P2 (PANSTARRS)	11	8	2011-06-10	2011-08-19	19.4	20.1	6.3	6.3	5.3	5.5
C/2011 Q4 (SWAN)	4	2	2012-02-25	2012-02-26	20.3	20.7	2.5	2.5	1.8	1.8
C/2011 R1 (McNaught)	9	6	2010-10-04	2010-12-13	19.5	20.7	7.0	7.5	6.3	6.6
C/2012 A1 (PANSTARRS)	4	4	2010-10-31	2011-01-01	20.4	21.2	10.0	10.2	9.3	10.3
C/2012 A2 (LINEAR)	3	2	2011-04-07	2011-11-18	18.9	20.9	4.7	6.1	5.1	5.2
C/2012 CH_{17} (MOSS)	3	2	2012-01-04	2012-01-06	19.2	20.1	3.7	3.7	3.1	3.2
C/2012 E1 (Hill)	54	26	2011-06-10	2012-05-29	19.4	20.5	7.5	7.8	6.7	7.4
C/2012 E3 (PANSTARRS)	1	1	2012-06-09	2012-06-09	20.5	20.5	5.1	5.1	4.8	4.8
C/2012 Q1 (Kowalski)	2	1	2011-10-01	2011-10-01	19.9	20.1	9.5	9.5	8.9	8.9

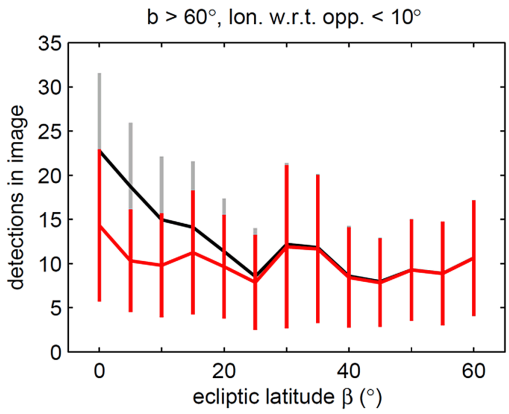


Figure 2.5: As in Figure 2.2, the vertical bars show the scatter (standard deviation) and the connected points are the mean values for the bins. Now added in red is the distribution of transients after exclusion of the ∼2 million known-object detections. The original distribution is included, in black, for comparison.

Of the 27,326 objects that were observed by *both* WISE and SDSS (satisfying the measurement error constraints mentioned above), 16,955 (62%) of these were also observed by PTF. A total of 624 of these WISE+SDSS objects were observed at least 10 times on at least one night in PTF, whence rotation curves can be estimated, while 625 of these WISE+SDSS objects have PTF observations in five or more phase-angle bins of width 3°, including opposition (0° –3°), whence phase functions can be estimated.

2.4 Unknown-object extraction

Exclusion of the ∼2 million known-object detections leaves ∼18 million transients remaining in our list of moving-object candidates. Figure 2.5 shows that the ecliptic distribution of transients per image has flattened out substantially. However, ignoring the scatter, the mean number of transients in the leftmost (lowest ecliptic latitude) bin remains the highest by more than two detections per image, suggesting the presence of significant unknown (i.e., undiscovered) small bodies in the data.

2.4.1 Previous and ongoing PTF small-body discovery work

In [Polishook et al. (2012)], a pilot study of rotation curve analysis and new-object discovery was undertaken using a few nights of $\sim$20 deg^2 high cadence ($\sim$20-minute-spaced) PTF data obtained in February 2010 at low ecliptic latitude ($|\beta| < 2.5°$). Using an original moving-object detection algorithm, they extracted 684 asteroids; of those which received provisional designations, three still qualify as PTF discoveries as of March 2013 (2010 CU_{247}, 2010 CL_{249} and 2010 CN_{249}). Though highly efficient on high cadence data, their tracklet-finding algorithm's limitations (e.g., single-night, single-CCD) renders it inapplicable to the vast majority of regular- (hour-to-days-) cadence PTF data.

A popular solution to this problem was already mentioned in Section 2.3, namely the use of kd-trees. A recently successful such kd-tree-based, detection-couplet-matching MOPS was used on the WISE data [Dailey et al. (2010)]. The WISE MOPS successfully extracted $\sim$2 million observations of $\sim$158,000 moving objects from the WISE data, including $\sim$34,000 new objects. A modified version of the WISE MOPS is under development for PTF at IPAC. As with the WISE MOPS, a key intent is the discovery of near-Earth objects, hence the PTF MOPS will need to accommodate relatively fast apparent motions (at least an order of magnitude faster than main-belt speeds). This poses considerable challenges, because PTF's cadences and false-positive detection rates are less accommodating than those of the space-based WISE survey. Though far from complete, the prototype PTF MOPS has successfully demonstrated that it can find tracklets spanning multiple nights and multiple fields of view, including at least two near-Earth objects, one of which was unknown (J. Bauer, personal communication).

As the PTF MOPS is still in development, for the purposes of this chapter we implement an original moving-object detection algorithm and run it on our residual $\sim$18 million-transient sample. The reader who wishes to skip over the details of the discovery algorithm should now go to Section 2.4.3.

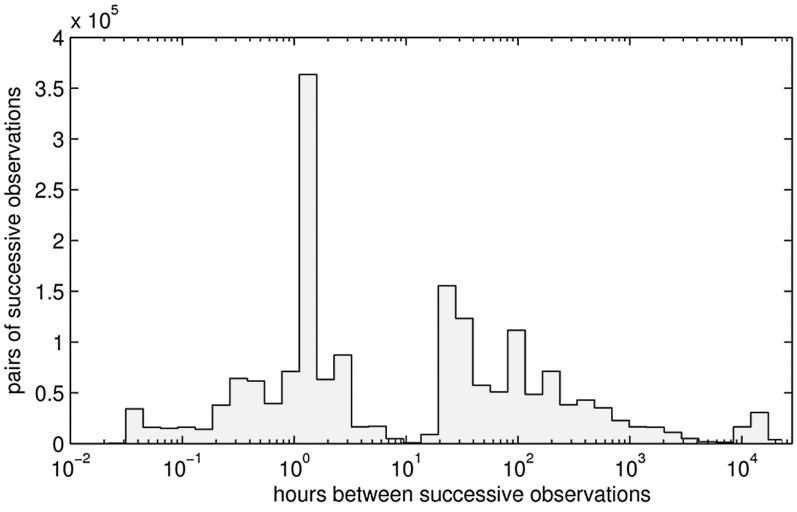


Figure 2.6: Time interval between consecutive observations of known objects in PTF. This distribution justifies the 48-hour upper limit we impose for tracklet finding, which also was chosen for computational expediency.

2.4.2 A custom discovery algorithm for main-belt objects

Because our intention is solely to supplement the main-belt comet search, we restrict apparent motions to those typical of main-belt objects (thereby easing the computational burden, but excluding faster NEOs and slower TNOs). This on-sky motion range is taken to be between 0.1 and 1.0 arcsec/minute.

Analysis of the known-object sample (Figure 2.6) shows that about half of all consecutive-observation pairs occur over a less than 12-hour (same night) interval, with a sharp peak at the one-hour spacing. Of the remaining (multi-night) consecutive-observation pairs, roughly half span less than 48 hours. Given these statistics, we prescribe 2 days as our maximum allowable timespan (between first and last observation) for a minimum three-point tracklet. As will be explained, multiple primary tracklets can be merged to produce a secondary tracklet greater than 2 days in total length, but the interval between any two consecutive points in the secondary tracklet still will not exceed 2 days. An imposed minimum time of 10 minutes between consecutive tracklet points ensures that the object has moved at least one arcsecond (for the minimum allowed speed), such that stationary transients (e.g., hostless-supernovae) are excluded.

Having specified time and velocity limits, the problem reduces to searching a double-cone-shaped volume, in three-dimensional time-plus-sky space surrounding each transient, to find sufficiently collinear past and future points. We modified our kd-tree implementation from Section 2.3 for this purpose. In particular, because PTF data were not collected on every consecutive night of the 41-months (due to weather, scheduling, etc.), the two-day upper limit we impose makes node-splitting along two-day (minimum) gaps in the data more natural and practical than simply splitting at median times, as was done in Section 2.3.1 and as is done generally for kd-trees.

An illustration of the tracklet-finding scheme (simplified to one spatial dimension) appears in Figure 2.7. For each transient, the kd-tree is used to rapidly find all other transients within its surrounding double-cone. Then, for every candidate past-plus-future pair of points, the two components of velocity and the distance residual of the middle transient from the candidate pair's predicted location (at the middle transient's epoch) are computed. Candidate past-plus-future pairs are then automatically discarded on the basis of the middle transient's distance residual with respect to them. To accommodate a limited amount of constant curvature, we use an adaptive criterion that is least stringent when the middle transient lies exactly at the midpoint between the past and future points, and becomes linearly more stringent as the middle transient nears one endpoint (approaching zero-tolerance at an endpoint). For candidate pairs spanning a single night or less, the maximum allowed middle-point residual is $1''$, while multi-night candidate pairs are allowed up to a $10''$ offset at the midpoint.

All remaining candidate past-plus-future pairs are then binned in two dimensions based on their two velocity components (R.A. and Dec. rates). Since the maximum allowed speed is 1 arcsec/minute, bins of 0.05 arcsec/minute between ± 1 arcsec/minute are used. If any single bin contains more candidate pairs than any other bin, all transients in all pairs in that bin, plus the middle transient, are automatically assigned a unique tracklet label. If more than one bin has the maximal number of pairs, the pair with the smallest midpoint residual is used. If any of these transients already has a tracklet label, all the others are instead assigned that existing label.

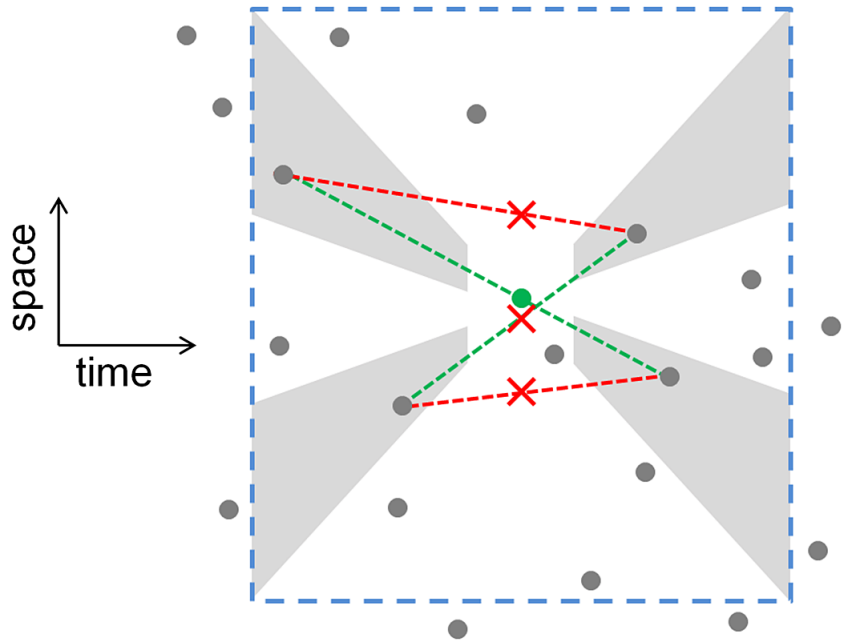


Figure 2.7: Schematic of our tracklet-finding algorithm, with only one spatial dimension (rather than the actual two) for clarity. Given all transients (gray dots), a kd-tree search rapidly finds those that are nearby in spacetime (within the blue dashed box, of 48-hour full-width) to the central, target transient (green dot). Minimal- and maximal-velocity bounds then define a subset of these (all dots lying in the gray shaded regions, in this case four). All possible past + future pairs are considered (in this case, four possible pairs). Pairs whose predicted midpoint position is sufficiently far from the target transient are immediately rejected (red dashed lines). Pairs with a sufficiently small residual (green dashed lines) then are binned in velocity space, and the velocity bin containing the most pairs is chosen. In this example, however, both non-empty velocity bins have only one pair, in which case the pair with the smallest midpoint residual is chosen.

Following this stage in the new-object discovery process, all tracklets found are screened rapidly by eye to eliminate false positives. The remaining tracklets are assigned a preliminary orbital solution using the orbit-fitting software *Find_Orb*[3] in batch mode. All orbital solutions are then used to re-search the transient data set for missed observations which could further refine the object's orbit. Because linear position extrapolation is replaced at this point by full orbital-solution-based ephemerides, the merging of tracklets across gaps in time longer than 48 hours is attempted in this last step.

2.4.3 Summary of objects discovered

We found 626 new objects which had a sufficient number of observations (at least two per night on at least two nights) to merit submission to the Minor Planet Center (MPC), whereupon they were assigned provisional designations. Four new comets were among the objects found by this moving-object search: 2009 KF_{37}, 2010 LN_{135}, 2012 KA_{51}, and C/2012 LP_{26} (see Table 2.4 in Section 2.6 for details). The first is a Jupiter-family comet and the latter three are long-period comets. As of March 2013, the first three still bear provisional asteroidal designations assigned by the MPC's automated procedures; the fourth, C/2012 LP_{26} (Palomar), was given its official cometary designation after follow-up observations were made in Februrary 2013 [Waszczak et al. (2013a)]. The cometary nature of these objects was initially noted on the basis of their orbital elements; an independent confirmation on the basis their measured extendedness appears in Section 2.6.

[3]The batch (non-interactive) Linux version of *Find_Orb* tries combinations of the Väisälä and Gauss orbit-determination methods on subsets of each tracklet in an attempt to converge on an orbit solution with minimized errors. For more information, see http://www.projectpluto.com/find_orb.htm.

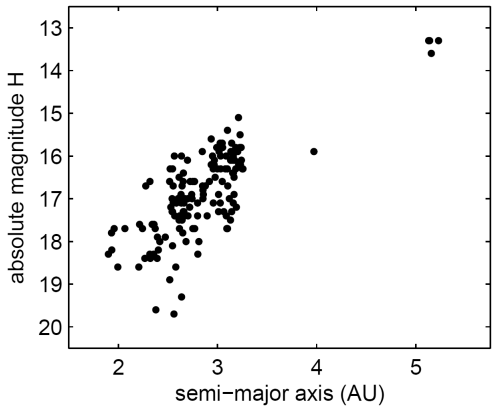


Figure 2.8: Of 622 asteroids discovered in multi-night PTF data through July 2012, the MPC has provided orbital solutions for 470 of these. Assuming typical albedos, the smallest ($H \sim 19.5$) objects correspond to ∼0.5-km diameters, while the $H \sim 13$ Trojans correspond to ∼10-km diameters.

2.5 Extended-object analysis: Approach

2.5.1 Definition of the extendedness parameter μ

To quantify the extendedness of a given small-body observation, we use the ratio of the object's total flux, within a flexible elliptical aperture (Kron 1980), to its maximum surface flux (i.e., the flux of the brightest pixel). Specifically, in terms of SExtractor output quantities, for each detection we define the quantity μ as `MU_MAX` − `MAG_AUTO` minus the median value of `MU_MAX` − `MAG_AUTO` for bright unsaturated stars on the image (note that the ratio of fluxes is equivalently the difference in magnitudes).

Unlike full-width at half maximum (FWHM), which is based on a one-dimensional symmetric (e.g., Gaussian) PSF model, μ is versatile as a metric in that it does not involve any assumption of symmetry (radial or otherwise). Note that in Section 2.2 we defined and excluded radiation hit candidates as those detections having $\mu < -1$. A negative μ means the object is more concentrated than bright stars on the image, while a positive μ means it is more extended. The error in μ, denoted σ_μ, is obtained by adding in quadrature the instrumental magnitude error `MAGERR_AUTO` and the 16th-to-84th percentile spread in `MU_MAX` − `MAG_AUTO` for the bright stars.

2.5.2 Systematic (non-cometary) variation in μ

The μ of a given small-body detection varies systematically with several known quantities, meaning that "extended" as defined by μ is not synonymous with "cometary".

Firstly, we must consider the apparent magnitude, since detections near the survey's limiting magnitude have a known bias [Ofek et al. (2012a)] in their instrumental magnitude (`MAG_AUTO`), which by definition affects the value of μ. [Ofek et al. (2012a)] note that use of the aperture magnitude `MAG_APER` rather than the adaptive Kron magnitude `MAG_AUTO` removes this bias, but unfortunately photometric zeropoints only exist presently for the latter in the PTF photometric database.

Secondly, the object's apparent motion on the sky during the 60-second exposure time must be considered, as such motion causes streaking to occur, which alters the flux distribution and hence also μ. It turns out that in $> 99\%$ of all observations in our sample (mostly main-belt objects), the on-sky motion is smaller than $1''$/minute. Given PTF's $1''$/pixel resolution, one might expect that the vast majority of objects are not drastically affected by streaking. Nevertheless, μ varies systematically with motion, as it does for apparent magnitude (see Section 2.5.4).

Thirdly, the photometric quality of an observation's host image, i.e., the seeing (median FWHM) and sky brightness, must be taken into account, since the median and spread of `MU_MAX` $-$ `MAG_AUTO` for bright stars on the image, and hence also μ, are influenced by such conditions.

A final measurable property affecting μ is the distance between the object's flux barycenter and the center of its brightest pixel. In terms of SExtractor quantities, this is computed as $((\texttt{XPEAK\_IMAGE} - \texttt{X\_IMAGE})^2 + (\texttt{YPEAK\_IMAGE} - \texttt{Y\_IMAGE})^2)^{1/2}$. In particular, if the barycenter lies near to the pixel edge ($\sim 0.5''$ from the pixel center), the majority of the flux will be nearly equally shared between two adjacent pixels. If it is near the pixel corner ($\sim 0.7''$ from the center), the flux will be distributed into four pixels (assuming a reasonably symmetric PSF and non-Poisson-noise-dominated signal). An object's position relative to the pixel grid is random, but the resulting spread in barycenter position does cause systematic variation in μ.

We can reasonably assume that some of these variations may be correlated. [Jedicke et al. (2002)] discusses systematic observable correlations of this kind, in the separate problem of debiasing sky-survey small-body data sets. Jedicke et al. also introduces a general formalism for representing survey detection systematics, which we now adapt in part to the specific problem of using μ to identify cometary activity.

2.5.3 Formalism for interpreting μ

Let the state vector $\vec{x}$ contain all orbital (e.g., semi-major axis, eccentricity) and physical (e.g., diameter, albedo) information about an asteroid. Given this $\vec{x}$ there exists a vector of observed quantities $\vec{o} = \vec{o}(\vec{x})$. Most of these observed quantities are a function of the large number of parameters defining the sky-survey (pointings, exposure time, optics, observatory site, data reduction, etc.). Included in $\vec{o}$ are the apparent magnitudes, on-sky motion, host-image seeing and sky brightness, barycenter-to-max-pixel distance, and also counts of how many total detections and how many unique nights the object is observed. In the above paragraphs we argued qualitatively that $\mu = \mu(\vec{o})$.

Now suppose that $\vec{x} \rightarrow \vec{x} + \delta\vec{x}$, where addition of the perturbing vector $\delta\vec{x}$ is equivalent to the asteroid exhibiting a cometary feature. For instance, $\delta\vec{x}$ could contain information on a mass-loss rate or the physical (3-dimensional) scale of a coma or tail. The resulting change in the observables is

$$\vec{o} \rightarrow \vec{o} + \delta\vec{o} \quad \text{where} \quad \delta\vec{o} = \frac{d\vec{o}}{d\vec{x}} \cdot \delta\vec{x} \tag{2.1}$$

The observation-perturbing vector $\delta\vec{o}$ could contribute to increased apparent magnitudes while leaving other observables such as sky position and apparent motion unchanged. To "model" the effect of cometary activity $\delta\vec{x}$ on the observables, e.g., as in [Sonnett et al. (2011)], is equivalent to finding (or inverting) the Jacobian $d\vec{o}/d\vec{x}$, though this is unnecessary for the present analysis. The resulting change in the scalar quantity μ is

$$\mu \rightarrow \mu + \delta\mu \quad \text{where} \quad \delta\mu = \nabla\mu \cdot \delta\vec{o} \tag{2.2}$$

Now suppose that some component of $\delta\vec{o}$ is (linearly) independent of $\vec{o}$, i.e., there exists some unit vector $\hat{\imath}$ such that $\hat{\imath} \cdot \delta\vec{o} = \delta o_i > 0$ while $\hat{\imath} \cdot \vec{o} = 0$. Another way of stating this is that there exists some observable o_i (the i^{th} component of $\vec{o}$), the value of which unambiguously discriminates whether the object is cometary or inert. An example would be the object's angular size on the sky. This need not be a known quantity; e.g., in the case of angular size one would need to employ careful PSF deconvolution to accurately measure it. The details of o_i do not matter, more important is its ability to affect μ, as described below.

Given the existence of this discriminating observable o_i, we can write

$$\delta\mu = \nabla\mu \cdot \delta\vec{o} = \delta\mu_{\text{sys}} + \frac{\delta\mu}{\delta o_i}\delta o_i \tag{2.3}$$

where the first term on the right side, $\delta\mu_{\text{sys}}$, represents systematic change in μ due to variation in known observables such as apparent magnitude and motion, and the second term represents a uniquely cometary contribution to μ. We assume that $\delta\mu/\delta o_i \neq 0$ in order for this reasoning to apply.

From our large sample of small-body observations, we are able to compare two objects, $\vec{o}$ and $\vec{o}\,'$, that have the same apparent magnitude, motion, seeing, etc. The computed $\delta\mu = \mu(\vec{o}) - \mu(\vec{o}\,')$ in such a case must have $\delta\mu_{\text{sys}} = 0$, meaning a result of $\delta\mu \neq 0$ would imply one of the objects is cometary. We can then use prior knowledge, e.g., that $\vec{o}$ is an inert object, to conclude $\vec{o}\,'$ is a cometary observation.

2.5.4 A model-μ to describe inert objects

We build upon this formalism by employing prior knowledge of the apparent scarcity of main-belt comets. That is, we hypothesize that the vast majority of known objects in our sample are in fact inert, or mapped to an equivalently inert set of observations $\vec{o}$ when subjected to the survey mapping $\vec{x} \rightarrow \vec{o}(\vec{x})$. This allows us to construct a

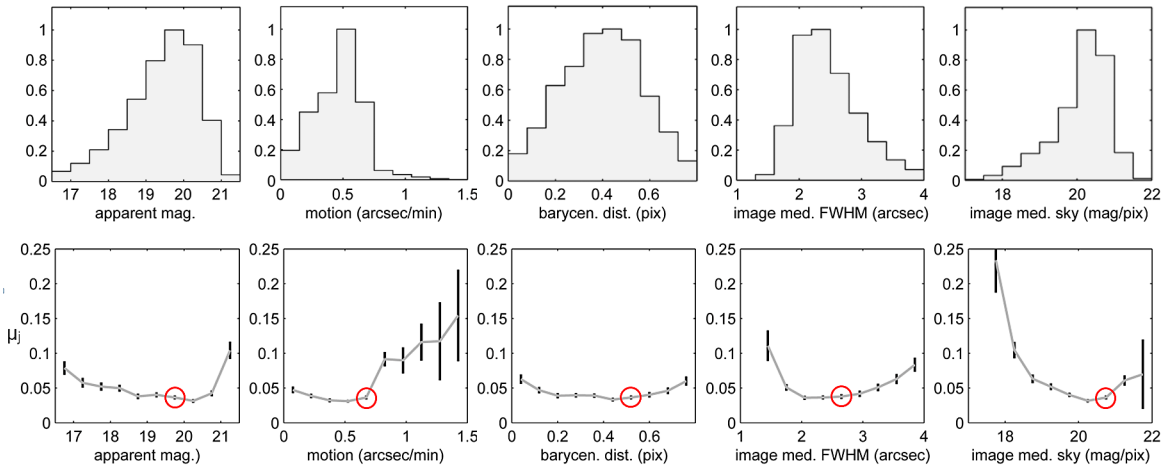


Figure 2.9: Distribution of PTF asteroid observations in the five parameters comprising the observable vector $\vec{o}$. Bottom row: Model-μ values (μ_j-values) plotted as a function of each parameter, while holding all other parameters constant; black bars show the error $\sigma_{\mu,j}$. The red circled point is the value at which the parameter is evaluated in the other plots. These plots only show a small slice of the much larger (and impossible to visualize) five-dimensional gridded function μ_j.

gridded model of μ for inert objects, denoted μ_j.

We first bin the data in a five-dimensional $\vec{o}$-space and then compute the error-weighted mean of μ in each bin. The j^{th} bin in this $\vec{o}$-space is defined as the five-dimensional box having corners $\vec{o}_j$ and $\vec{o}_j + \Delta\vec{o}$. The model value μ_j in this j^{th} bin is found by summing over all observations in that bin:

$$\mu_j = \sigma_{\mu,j}^2 \sum_{\vec{o}\in[\vec{o}_j,\vec{o}_j+\Delta\vec{o}]} \frac{\mu(\vec{o})}{\sigma_\mu(\vec{o})^2} \tag{2.4}$$

where the scatter (variance) in the j^{th} bin is

$$\sigma_{\mu,j}^2 = \left(\sum_{\vec{o}\in[\vec{o}_j,\vec{o}_j+\Delta\vec{o}]} \frac{1}{\sigma_\mu(\vec{o})^2} \right)^{-1} \tag{2.5}$$

and the individual observation errors σ_μ are computed as described in Section 2.5.1. We exclude known comets from all bin computations, even though their effect on the mean would likely be negligible given their small population relative to that of asteroids.

The histograms in Figure 2.9 show the range of values for the five components of $\vec{o}$, each of which is sampled in 10 bins. The five-dimensional $\vec{o}$-space considered thus has 10^5 bins. However, given the centrally-concentrated distributions of each observable, only a fraction ($\sim$40%) of these bins actually contain data points. Some bins ($\sim$9%) only include a single data point; these data cannot be corrected using this μ_j model, but their content represent $< 1\%$ of the data. Lastly, $\sim$7% of the data lie outside one or more of these observable ranges, and hence also cannot be tested using the model. Most of these excluded data are either low quality (seeing $> 4''$) or bright objects (> 16.5 mag). Of the 175 previously known comets (see Section 2.3) plus 4 new (see Section 2.4) comets we found in PTF, 115 of these (64%, mostly the dimmer ones) lie in these observable ranges and hence can be tested with the model.

2.5.5 Defining a visually-screenable sample

For each of the $\sim$2 million small-body observations in our data set, we use the inert model μ_j to define the corrected extendedness as a "μ-excess":

$$\delta\mu = \mu - \mu_j \tag{2.6}$$

and an uncertainty:

$$\sigma = \sqrt{\sigma_\mu^2 + \sigma_{\mu,j}^2} \tag{2.7}$$

For each of the $\sim$220,000 unique objects in our data set, we sum over all observations of that object to define

$$\overline{\delta\mu} = \langle\delta\mu\rangle^2 \quad \sum_{\substack{\text{object's}\\ \text{observ-}\\ \text{ations}}} \frac{\delta\mu}{\sigma^2} \tag{2.8}$$

$$\langle \delta\mu \rangle = \left(\sum_{\substack{\text{object's} \\ \text{observ-} \\ \text{ations}}} \frac{1}{\sigma^2} \right)^{-1/2} \tag{2.9}$$

These two quantities, $\overline{\delta\mu}$ and $\langle \delta\mu \rangle$, are useful for screening for objects which appear cometary in most observations. If an object is observed frequently while inactive but sparsely while active, $\overline{\delta\mu}$ and $\langle \delta\mu \rangle$ are less useful. As noted in Section 2.3.3, high cadence data are uncommon in our sample, alleviating this problem (see also Figure 2.15 in Section 2.7 for commentary on orbital coverage).

To select the sample to be screened by eye for cometary activity, we use the quantity $\overline{\delta\mu} - \langle \delta\mu \rangle$. In the case of normally-distributed data, the probability that this quantity is positive is $1 - \mathrm{erf}(1) \approx 0.16$. As shown in Figure 2.10, the fraction of objects with $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$ is actually 0.007 (1,577 objects), much smaller than the Gaussian-predicted 0.16. This likely results from overestimated (i.e., larger-than-Gaussian) $\langle \delta\mu \rangle$ values caused by outliers in the data. However, of the 115 testable comets in our data (111 known plus 4 new—see Table 2.5 for further explanation), 76 of these (66%) have $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$. That is, a randomly chosen known comet from our sample is ∼100 times more likely to have $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$ than a randomly chosen asteroid from our sample, suggesting the criterion $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$ is a robust indicator of cometary activity.

The fact that only 66% of the 115 comets in our testable sample satisfy $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$ means that, if one assumes main-belt comets share the same extendedness distribution as all comets, then our detection method is only 66% efficient. Sufficiently weak and or unresolved (very distant) activity inevitably causes the lower and negative values of $\overline{\delta\mu} - \langle \delta\mu \rangle$.

Given the specific goal to detect main-belt objects that are at least as active as the known candidate MBCs, we consider the value of $\overline{\delta\mu} - \langle \delta\mu \rangle$ for the known candidate MBCs in our sample. P/2010 R2 (La Sagra) has $\overline{\delta\mu} - \langle \delta\mu \rangle = 0.474$ (from 34 observations made on 21 nights). P/2006 VW_{139} has $\overline{\delta\mu} - \langle \delta\mu \rangle = 0.231$ (from 5 observations made on 3 nights). Hence, the $\overline{\delta\mu} - \langle \delta\mu \rangle > 0$ criterion is more than

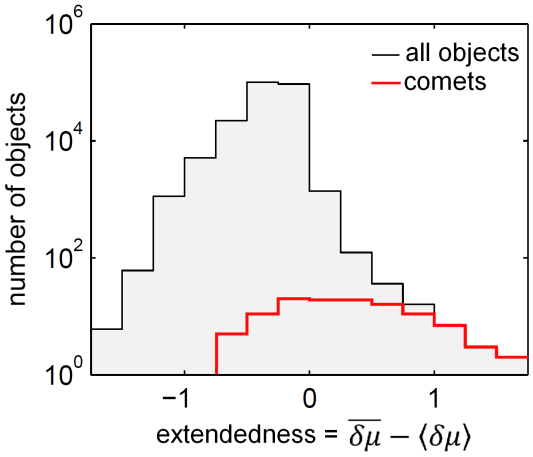


Figure 2.10: For inclusion in the screening sample, an object's mean extendedness value, $\overline{\delta\mu}$, must exceed zero by more than one-sigma, $\langle\delta\mu\rangle$, which in this histogram is true for all objects to the right of zero. The 111 known (and testable) comets plus the 4 new comets (see Section 2.4.3) are plotted in red; 76 out of these 115 fall to the right of zero, meaning our method is 76/115 = 66% efficient at detecting the known plus new comets comprising our sample. Both of the MBCs in our sample fall to the right of zero, implying that we are 100% efficient at detecting objects at least as extended as these known MBCs.

sufficient (formally, 100% efficient) for detecting extendedness at the level of these known, kilometer-scale candidate MBCs. Note however that we do *not* claim 100% detection efficiency with respect to objects of similar *magnitude* as these candidate MBCs; see Figure 2.11 for a consideration of efficiency as a function of apparent magnitude.

2.6 Extended-object analysis: Results

A total of 1,949 observations (those having $\delta\mu - \sigma > 0$) of 1,577 known and newly discovered objects satisfying $\overline{\delta\mu} - \langle\delta\mu\rangle > 0$ were inspected visually to identify either contamination from image artifacts or true cometary features. For each detection this involved viewing a $2' \times 2'$ cutout of the image, with contrast stretched from -0.5σ to $+7\sigma$ relative to the median pixel value (where $\sigma = \sqrt{\text{median}}$). This image was also flashed with the best available image of the same field taken on a different night ("best" meaning dimmest limiting magnitude), to allow for rapid contaminant

identification.

With the exception of two objects (described below), virtually all of these observations were clearly contaminated by either a faint or extended nearby background source, CCD artifacts or optical artifacts (including ghosts and smearing effects). In principle these observations should have been removed from the list of transients by the filtering process described in Section 2.2.2, however some residual contamination was inevitable.

The screening process did however reveal cometary activity on two non-main-belt objects previously labeled as asteroids: 2010 KG_{43} and 2011 CR_{42}, which had $\overline{\delta\mu} - \langle\delta\mu\rangle$ values of 0.2 and 1.1, respectively (Figure 2.12). Note that taking these two objects into account improves our efficiency slightly to $(76+2)/(115+2) = 67\%$.

In addition to the two known candidate MBCs (Figure 2.13)—which were among the 76 comets already noted to have passed the screening procedure—this process also confirmed the extendedness of three of the four comets discovered in PTF as moving-objects (2009 KF_{37}, 2010 LN_{135}, and 2012 KA_{51}) as described in Section 2.4. These comets had $\overline{\delta\mu} - \langle\delta\mu\rangle$ values of 0.33, 0.29, and 0.58, respectively. The procedure did *not* identify the fourth new comet discovery, C/2012 LP_{26} (Palomar), as an extended object, suggesting that it was unresolved.

2.6.1 A new quasi-Hilda comet: 2011 CR_{42}

The object 2011 CR_{42}, discovered on 2011-Feb-10 by the Catalina sky-survey [Drake et al. (2009)], has an uncommon orbit ($a = 3.51$ AU, $e = 0.28$ and $i = 8.46°$). Six PTF g'-band observations made between 2011-Mar-05 and Mar-06 [Waszczak et al. (2011)] all show a coma-like appearance but no tail. The object was 2.92 AU from the Sun and approaching perihelion ($q = 2.53$ AU on 2011-Nov-29). Based on its orbit and using IAU phase-function parameters [Bowell et al. (1989)] $H = 13.0$ and $G = 0.15$, 2011 CR_{42} should have been easily observed at heliocentric distances 3.8 AU and 3.1 AU in 2010-Feb and 2010-Dec PTF data, with predicted magnitudes of 19.2 and 18.8 mag, respectively. Upon inspection of these earlier im-

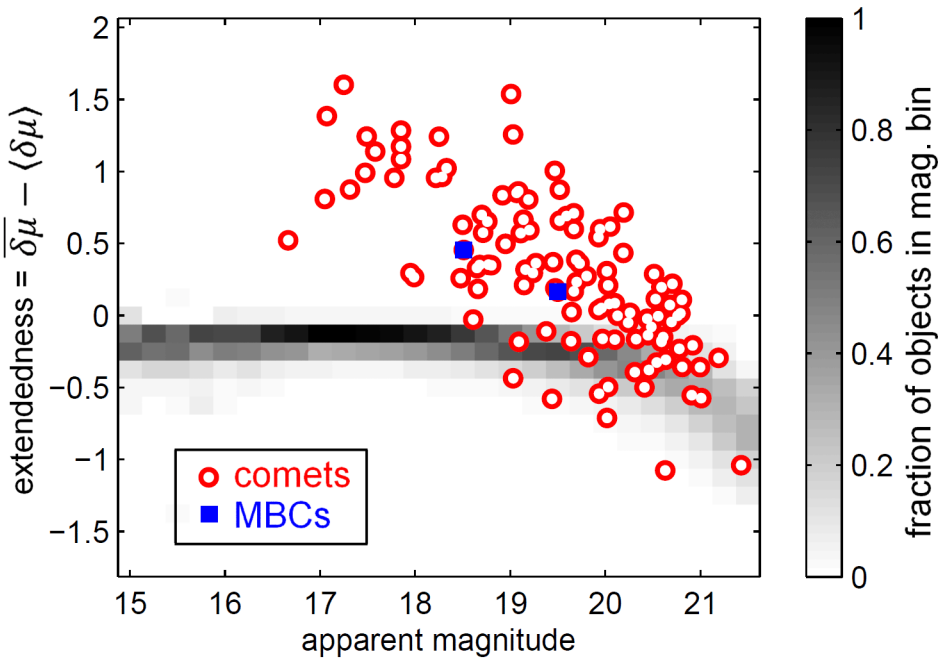


Figure 2.11: Like Figure 2.10, but now with each object's median apparent magnitude plotted as well. All small bodies in the sample are represented in the 2D histogram (normalized with respect to each magnitude bin), while the known comets are overplotted as red-white circles, the two MBCs as blue squares. Again, about two-thirds (66%) of the comets lie above zero. This plot suggests that the completeness is expressible as a function of apparent magnitude. That is, $C > 66\%$ for bright comets and $C < 66\%$ for dim comets (approaching zero for > 21 mag), while on average $C = 66\%$. The exact magnitude dependence is sensitive to bin size and is not explored quantitatively here.

Table 2.4: Comets discoveries made in PTF in the course of this work

name	q (AU)	e	i (deg)	$T_{\rm peri}$	dates in PTF	type*	co-discoverer	Reference†
2009 KF_{37}	2.59	0.34	11.2	2009-Aug	2009-May to 2009-Jul	JF	—	MPS 434214
2010 KG_{43}	2.89	0.49	13.5	2010-Jul	2010-Aug to 2010-Sep	JF	WISE (discovered orbit)	MPS 434201
2010 LN_{135}	1.74	1.00	64.3	2011-May	2010-Jun to 2010-Jul	LP	—	MPS 439624
2011 CR_{42}	2.53	0.28	8.5	2011-Nov	2011-Mar	JF	Catalina (discovered orbit)	CBET 2823
2012 KA_{51}	4.95	1.00	70.6	2011-Nov	2012-May	LP	—	MPS 434214
C/2012 LP_{26} (Palomar)	6.53	1.00	25.4	2015-Aug	2012-Jun to 2012-Jul	LP	Spacewatch (discovered coma)	CBET 3408

*JF = Jupiter-family; LP = long-period †MPS = *Minor Planet Circulars Supplement*; CBET = *Central Bureau Electronic Telegram*

ages, no object was found within 200″ of the predicted position. Its absence in these images further suggests cometary activity.

Like the MBCs and unlike most Jupiter-family comets, 2011 CR_{42}'s Tisserand parameter [Murray and Dermott (1999)] with respect to Jupiter ($T_{\rm Jup} = 3.042$) is greater than 3. While the criterion $T_{\rm Jup} > 3$ is often used to discriminate MBCs from other comets, we note that MBCs more precisely have $T_{\rm Jup} > 3.1$. About half of the ∼20 *quasi-Hilda comets* (QHCs, [Toth (2006)] and refs. therein) have $3 < T_{\rm Jup} < 3.1$, as does 2011 CR_{42}. Three-body (Sun + Jupiter) interactions tend to keep $T_{\rm Jup}$ approximately constant (this is akin to energy conservation). Such interactions nonetheless can chaotically evolve the orbits of QHCs. Their orbits may settle in the stable 3:2 mean-motion (Hilda) resonance with Jupiter at 4 AU, wander to a high-eccentricity Encke-type orbit, or scatter out to (or in from) the outer Solar System. Main-belt orbits, however, are inaccessible to these comets under these $T_{\rm Jup}$-conserving three-body interactions.

To verify this behavior, we used the hybrid symplectic integrator MERCURY [Chambers (1999)] to evolve 2011 CR_{42}'s orbit forward and backward in time to an extent of 10^4 years. For initial conditions we tested all combinations of 2011 CR_{42}'s known orbital elements plus or minus the reported error in each (a total of $3^6 = 729$ runs in each direction of time). We did not include non-gravitational (cometary) forces in these integrations, as it was assumed that this object's relatively large perihelion distance would render these forces negligible. In ∼25% of the runs, the object scattered out to (or in from) the outer solar System in less than the 10^4 year duration of the run. In the remainder of the runs, its orbit tended to osculate about the stable 3:2 mean motion Jupiter resonance at 4 AU. These results strongly suggest that 2011 CR_{42} is associated with the Hilda family of objects belonging to this resonance, and thus likely is a quasi-Hilda comet.

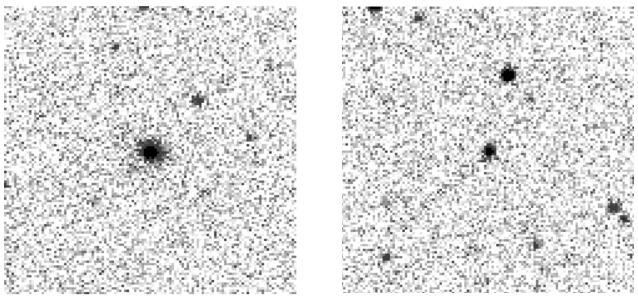


Figure 2.12: Known asteroidally-designated objects whose cometary activity was discovered in PTF in the course of this work. Each image is $2' \times 2'$ (pixel scale $1.01''$). Left: 2011 CR_{42} in g'-band on 2011-03-06. No tail is discernible, but the object's FWHM is twice that of nearby stars. Right: 2010 KG_{43} in R-band on 2010-09-08. A $\sim 1'$-long tail is discernible extending toward the lower left corner of the image.

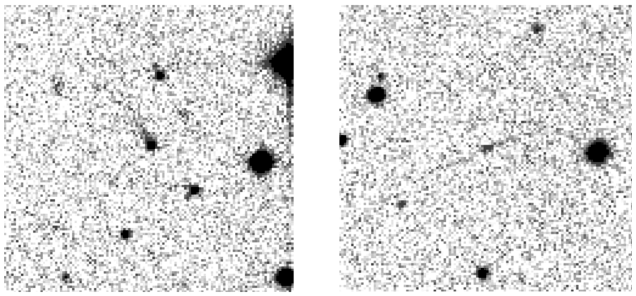


Figure 2.13: Known candidate main-belt comets in PTF. Each image is $2' \times 2'$ (pixel scale $1.01''$). Left: P/2010 R2 (La Sagra) in R-band ($R \sim 18.5$ mag) on 2010-08-19, with its tail extending towards the top left. Right: P/2006 VW_{139} in g'-band ($g' \sim 20$ mag) on 2011-12-21, with its two oppositely oriented tails barely discernible by eye.

Table 2.5: Summary of the PTF comet sample. JF = Jupiter-family; LP = long-period; MB = main-belt. "Observed" means found by the search algorithms of Sections 3 or 4; "model-μ tested" means it lies in the observable ranges shown in Figure 2.9 (e.g., excludes bright comets), and $\overline{\delta\mu} - \langle\delta\mu\rangle > 0$ means positively detected as extended. Objects 2010 KG_{43} and 2011 CR_{42} are counted in all rows as PTF-discovered JFCs, even though they were not included in the $76/115 = 66\%$ efficiency calculation of Section 2.5.5 (since they were not discovered until Section 2.6).

	previously known			PTF discovered			
	JF	LP	MB	JF	LP	MB	total
observed	108	65	2	3	3	0	181
model-μ tested	71	38	2	3	3	0	117
$\overline{\delta\mu} - \langle\delta\mu\rangle > 0$	44	27	2	3	2	0	78

2.7 Statistical interpretation

2.7.1 Bayesian formalism

Following the approach of [Sonnett et al. (2011)] and borrowing some of their notation, we apply a Bayesian formalism to our survey results to estimate an upper limit for the fraction f of objects (having $D > 1$ km) which are active MBCs at the time of observation. The prior probability distribution on f is chosen to be a log-constant function:

$$P(f) = -\frac{1}{f \log f_{\min}} \quad \text{for} \quad f_{\min} < f < 1 \tag{2.10}$$

This prior is justified since we know f is "small", but not to order-of-magnitude precision. The minimum value $f_{\min} > 0$ is allowed to be arbitrarily small, since the integral of $P(f)$ is always unity:

$$\int_{f_{\min}}^{1} P(f)\, df = 1 \tag{2.11}$$

Let N be the number of objects in a given sample, n the number of active MBCs positively detected in that sample, and C the completeness or efficiency of our MBC-detection scheme. In Section 2.5.5 we discussed how $C = 0.66$ if defining completeness with respect to the extendedness distribution of the 115 known comets on which we tested our detection method. Relative to objects at least as extended as the two known candidate MBCs we tested, however, we can take $C = 1$ (100% efficiency), since both of the MBCs observed were robustly flagged by our screening process.

The likelihood probability distribution function for a general sample S is formally a binomial distribution, but because the samples we will be considering are very large ($N \gg 1$), the likelihood function is well-approximated as a Poisson distribution:

$$\begin{aligned} P(S|f) &= \frac{N!}{n!(N-n)!}(Cf)^n(1-Cf)^{N-n} \\ &\approx \frac{(NCf)^n}{n!}\exp(-NCf) \end{aligned} \tag{2.12}$$

Bayes' Theorem then gives the formula for the posterior probability distribution on f given our results:

$$P(f|S) = \frac{P(S|f)P(f)}{\int_{f_{\min}}^{1} P(S|f)P(f)\ df} \propto f^{n-1} \exp(-NCf) \tag{2.13}$$

The constant of proportionality (not shown) involves incomplete gamma functions[4], and is well-defined and finite (including in the limit $f_{\min} \to 0$).

Finally, we can compute the 95% confidence upper limit f_{95} by solving the implicit equation

$$\int_{0}^{f_{95}} P(f|S)\ df = 0.95 \tag{2.14}$$

2.7.2 Active MBCs in the *entire* main-belt

We first take the sample S to be representative of *all* main-belt asteroids, which in our survey amounted to $N = 2.2 \times 10^5$ observed objects and $n = 2$ detected MBCs. Equation (2.14) yields 95% confidence upper limits for f of 33×10^{-6} and 22×10^{-6}, for efficiency values of C of 0.66 and 1.0, respectively. Figure 2.14 depicts the probability distributions for each case.

We note that although these results are based on the positive detection of only two candidate MBCs, the reader need not be skeptical on the basis of "small number statistics", since the Possionian posterior (Equation 2.13 and Figure 2.14) formally accounts for "small number statistics" through its functional dependence on n. Even if we had detected no MBCs at all—in which case n would be zero (as was the case in [Sonnett et al. (2011)])—the posterior would still be well-defined; the 95%-confident upper limit would naturally be larger to reflect the greater uncertainty.

Our discussion has so far only considered the *fraction* of active MBCs, rather than

[4]The *incomplete gamma function* is defined as

$$\Gamma(n, N) = \int_{N}^{\infty} t^{n-1} \exp(-t)\ dt$$

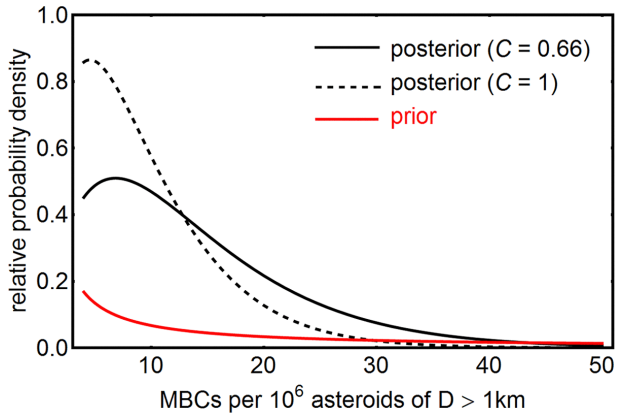


Figure 2.14: Probability distributions for estimating the fraction of main-belt comets, based on the results of our sample screening. The $C = 0.66$ case reflects our efficiency with respect to the extendedness distribution of all known comets, while the $C = 1$ case applies to extendedness levels at least as high as the two known candidate MBCs in our sample. In this plot $f_{\rm min}$ was set to 4×10^{-6}, to facilitate visual comparison of the normalized prior with the normalized posteriors.

their *total number*. This is because an estimate of the total underlying number of main-belt objects (down to $D \sim 1$ km) must first be quoted from a properly-debiased survey. A widely-cited example is [Jedicke and Metcalfe (1998)], who applied a debiasing analysis to the Spacewatch survey and concluded that there are of order 10^6 main-belt asteroids (down to $D \sim 1$ km)[5]. Since the MBC-fraction estimates in the above paragraphs are conveniently given in units of per million main-belt asteroids, we directly estimate the upper limit on the total number of active MBCs in the true underlying $D > 1$ km population (again to 95% confidence) to be between 33 and 22, depending on the efficiency factor C (0.66 or 1.0).

2.7.3 Active MBCs in the *outer* main-belt

Of the seven candidate MBCs listed in Table 2.1, all except 259P/Garradd have semi-major axes between 3.0 AU and 3.3 AU, corresponding approximately to the 9:4 and 2:1 Jupiter resonances (Kirkwood gaps). This semi-major axis constraint is satisfied by 123,366 (∼20%) of the known objects as of August 2012, of which 47,450 (38%)

[5]More recent survey results will eventually test/verify this result, e.g. the WISE sample has already produced a raw size-frequency distribution [Masiero et al. (2011)], the debiased form of which will be of great value.

are included in the PTF sample.

Reapplying equations (10)–(14) except now using $N = 47,450$ (while $n = 2$ remains unchanged), we find 95%-confidence upper limits of 160 and 110 active MBCs per million outer main-belt asteroids with $3.0 < (a/\mathrm{AU}) < 3.3$ and $D > 1$ km, for detection efficiencies of $C = 0.66$ and $C = 1.0$, respectively.

Although only ∼20% of the *known* main-belt asteroids lie in this orbital range, the debiased semi-major axis distribution presented in [Jedicke and Metcalfe (1998)] predicts that ∼30% of all main-belt objects (of $D > 1$ km) lie in this outer region. The discrepancy is due to the fact that these objects are more difficult to detect, since they are further away and tend to have lower albedos (this lower detection efficiency is evident for instance in the WISE sample shown in Figure 2.3). Assuming 300,000 objects actually comprise this debiased outer main-belt region, the inferred 95% confidence upper limit on the total number of active MBCs existing in this region is ∼50 (for $C = 0.66$) and ∼30 (for $C = 1.0$).

2.7.4 Active MBCs in the *low-inclination* outer main-belt

Four out of the seven candidate MBCs in Table 2.1 have orbital inclinations of $i < 5°$. Combined with the semi-major-axis constraint $3.0 < (a/\mathrm{AU}) < 3.3$, this associates them with (or close to) the Themis asteroid family. There are 25,069 objects in the small-body list we used which satisfy this combined constraint on a and i (∼4% of the known main-belt), of which 8,451 (34%) are included in the PTF sample.

Again reapplying equations (10)–(14), we now use $N = 8,451$ and $n = 1$, where the new value for n reflects the fact that P/2006 VW_{139} satisfies this i-criterion while P/2010 R2 (La Sagra) does not. We find 95%-confidence upper limits of 540 and 360 active MBCs per million low-inclination, outer main-belt asteroids, for detection efficiencies of $C = 0.66$ and $C = 1.0$, respectively.

Once again, [Jedicke and Metcalfe (1998)] offer estimates of the debiased number of objects in the underlying population of interest: for outer main-belt asteroids, they found that ∼20% of the debiased objects had $i < 5°$. Hence, assuming there are

60,000 objects (of $D > 1$ km) in the actual low-i outer main-belt population satisfying these a and i constraints, the resulting upper limit estimates for the total number of active MBCs it contains is ∼30 (for $C = 0.66$) and ∼20 (for $C = 1.0$).

2.7.5 Active MBCs among low-i outer main-belt objects observed *near perihelion* ($-45^\circ < \nu < 45^\circ$)

Of the 8,451 low-inclination outer main-belt objects observed by PTF (see Section 2.7.4), 5,202 were observed in the orbital quadrant centered on perihelion (in terms of true orbital anomaly ν, this quadrant is $-45^\circ < \nu < 45^\circ$). We consider this constraint given that all known candidate MBCs (Table 2.1) have shown activity near perihelion. Now using $N = 5,202$ and $n = 1$ (here again $n = 1$ represents P/2006 VW_{139}), we find 95%-confidence upper limits of 880 and 570 active MBCs per million low-inclination, outer main-belt asteroids observed by PTF near perihelion, for detection efficiencies of $C = 0.66$ and $C = 1.0$, respectively.

We caution that, unlike the previous subsets (which were defined solely by orbital elements), the population to which these statistics apply is less well-defined. In particular, the bias for detection near perihelion (Figure 2.15), due in part to the $(r\Delta)^{-2}$ dependence in the reflected sunlight, is more pronounced for smaller, lower-albedo, higher eccentricity objects. Hence, naively imposing a constraint on true anomaly ν implictly introduces selection biases in D, p_V and e. Moreover, these implicit biases depend on the sensitivity of the PTF survey in a more nuanced manner, invalidating the simple $D \gtrsim 1$-km lower limit we have quoted generally in this chapter. Nonetheless, these parameters (D and p_V) are important enough to merit individual treatment, as detailed below.

2.7.6 Active MBCs in the sub-5 km diameter population

Yet another constraint that well-encompasses the known MBC candidates of Table 2.1 is a diameter $D < 5$ km (corresponding to approximately $H > 15$ mag for albedo $p_V = 0.07$). Applying this constraint decreases the number of PTF-sampled objects N

by 28%, 45% and 41% for the entire main-belt, outer main-belt, and low-i outer main-belt, respectively. These smaller sample sizes result in slightly higher 95%-confidence upper limits for the fraction of active MBCs: 30–45, 180–280, and 610–920 per 10^6 objects having $5 > (D/\mathrm{km}) > 1$ in the entire main-belt, outer main-belt, and low-i outer main-belt, respectively (the ranges corresponding to the two values of the efficiency factor C).

[Jedicke and Metcalfe (1998)] found that the debiased differential number distribution as a function of absolute magnitude H is $\sim 10^{\alpha H}$, where $\alpha \approx 0.35$. The resulting cumulative number distribution (i.e., the number of asteroids brighter than absolute magnitude H) is $\sim 10^{\alpha H}/\alpha \ln(10)$. Using $H = 17$ in this expression gives the predicted 10^6 asteroids having $D > 1$ km. The fraction of these objects in the range $15 < H < 17$ is therefore $1 - 10^{\alpha(15-17)} \approx 80\%$. Scaling the debiased populations discussed above by this factor and using the new limits from the preceding paragraph gives new upper limits on the *total* number of active MBCs existing in the three regions: ∼24–36, ∼40–70, and ∼30–45 in the entire main-belt, outer main-belt, and low-i outer main-belt, respectively.

2.7.7 Active MBCs among low-albedo (WISE-sampled) objects

[Bauer et al. (2012)] analyzed WISE observations of five of the active-main-belt objects listed in Table 2.1. By fitting thermal models to the observations, they found that all of these objects had visible albedos of $p_V < 0.1$. As shown in Figure 2.3 and described in Section 2.3.4, about half of the asteroids which were observed by WISE also appear in the PTF sample; in particular there were $N = 32,452$ low-albedo ($p_V < 0.1$) objects observed by both surveys. Included in the [Bauer et al. (2012)] sample was PTF-observed candidate MBC P/2010 R2 (La Sagra), whose fitted albedo of $p_V = 0.01 \pm 0.01$ implies we can take $n = 1$ (one positive active MBC-detection) in the low-albedo WISE/PTF sample.

Following the 95% confidence upper limit computation method of the previous

sections, we derive upper limits of 90–140 active MBCs per 10^6 low-albedo ($p_V < 0.1$) asteroids. As mentioned earlier, the full-debiasing of the WISE albedo distribution [Masiero et al. (2011)] will eventually allow us to convert this upper limit on the *fraction* of active MBCs among low-albedo asteroids into an upper limit on their *total number*, just as [Jedicke and Metcalfe (1998)] has allowed us to do for orbital and size distributions.

2.7.8 Active MBCs among C-type (SDSS-sampled) objects

The MBC candidate P/2006 VW_{139} was observed serendipitously by SDSS on two nights in September 2000. While one of the nights was not photometric in g-band, the other night provided reliable g, r, i multi-color data on this object, yielding a principal component color $a^* = -0.14 \pm 0.05$. Because it has $a^* < 0$, this suggests P/2006 VW_{139} is a carbonaceous (C-type) object[6].

Figure 2.3 depicts the overlap of the SDSS-observed sample with PTF, which includes $N = 24,631$ C-type ($a^* < 0$) objects. Taking $n = 1$, we derive 95% confident upper-limits of 120–190 active MBCs per 10^6 C-type asteroids (where again the range corresponds to $C = 0.66$–1.0).

2.8 Conclusion

2.8.1 Summary

Using original kd-tree-based software and stringent quality filters, we have harvested observations of ∼40% (∼220,000) of the known solar system small bodies and 626 new objects (622 asteroids and 4 comets) from the first 41 months of PTF survey data (March 2009 through July 2012). This sample is untargeted with respect to the orbital elements of known small bodies (but not necessarily the *true* underlying population), down to ∼1-km diameter-sized objects. Most (∼90%) of the objects are

[6]This taxonomic classification for P/2006 VW_{139} has been confirmed spectroscopically by [Hsieh et al. (2012b)] and [Licandro et al. (2013)].

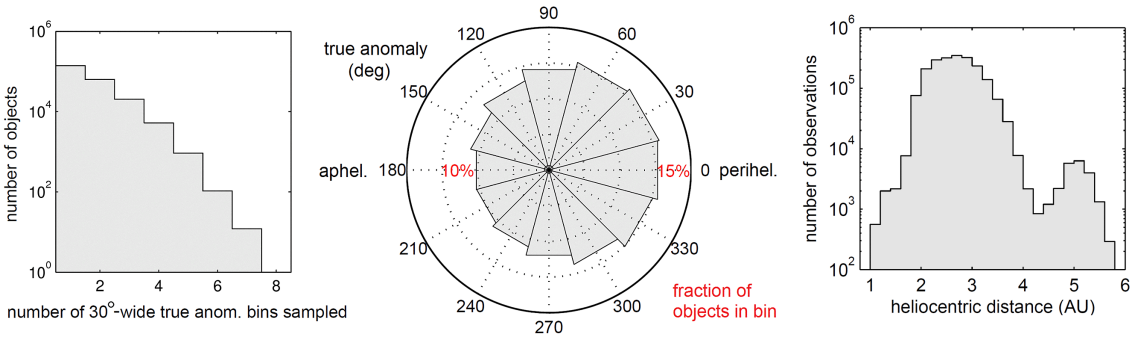


Figure 2.15: Summary of true-anomaly and heliocentric distance coverage of known small bodies in PTF (see Figure 2.3 for other orbital statistics of this sample), from March 2009 through July 2012. Most (∼90%) of objects have only been sampled in at most two 30°-wide true anomaly bins, i.e., less than 1/6 of the orbit. In the left histogram objects are only counted once, but in the rose diagram (middle), each object is counted once for each bin it in which it is sampled (hence the fraction values reported for all twelve bins sum to more than 100%). Although all objects spend more time around aphelion, most only fall above the survey detection limit near perihelion, hence there is a slightly larger fraction of objects observed near perihelion.

observed on less than ∼10 distinct nights, and ∼90% are observed over less than 1/6 of their orbit, allowing us to characterize this sample predominantly as a "snapshot" of objects in select regions of their orbits.

We have introduced a metric for quantifying the extendedness of a small-body in an image, and have corrected this metric, on a per-observation basis, for systematic variation due to observables such as apparent magnitude, on-sky motion and pixel-grid alignment. In this metric, an extendedness of zero describes stellar-like (asteroidal) objects, whereas a positive value indicates potentially cometary extendedness.

We defined a sample for visual screening consisting of all objects whose mean extendedness value is greater than zero by at least one-sigma. This screening sample consisted of ∼1,500 unique objects, 76 (out of 115) comets, and two known candidate active MBCs, P/2010 R2 (La Sagra) and P/2006 VW_{139}, which upon inspection appear active and visibly extended in the images. Of the ∼1,500 objects screened, we found evidence for activity on two known (non-main-belt) asteroidally-designated objects, 2010 KG_{43} and 2011 CR_{42}, and confirmed activity on the three out of the

four (non-main-belt) comets that our moving-object algorithm discovered.

Given these results, using a log-constant prior we infer with 95% confidence an upper limit of < 33 active MBCs per 10^6 main-belt asteroids for a $C = 0.66$ detection efficiency with respect to the extendedness distribution of known comets, and < 22 active MBCs per 10^6 main-belt asteroids for a 100% efficiency with respect to objects at least as extended as the known candidate active MBCs in our sample.

2.8.2 Comparison to previous work

Our inferred 95% confidence upper limit of at most ∼30 active MBCs per 10^6 main-belt asteroids of $D > 1$ km is comparable but slightly lower than that of [Gilbert and Wiegert (2010)], who estimated 40 ± 18 active MBCs per 10^6 main-belt asteroids, also for $D > 1$ km, from visual inspection of a similarly untargeted sample of ∼25,000 objects from the *Canada-France-Hawaii Telescope Legacy Survey* ([Gilbert and Wiegert (2010)]). That result was based on the detection of a single unknown comet in their sample, which was never actually confirmed to be a main-belt object due to lack of follow-up observations. Even before taking into account our order-of-magnitude larger sample size, we note that, in contrast to their results, our limits are based on positive MBC detections and use detection efficiencies estimated from observations of ∼100 known comets.

The result of [Sonnett et al. (2011)] was a much larger upper limit of ∼3,000 MBCs per 10^6 main-belt asteroids (to 90% confidence), albeit applicable to the smaller limiting diameter of ∼0.5 km. Their smaller sample size of 924 objects is certainly the cause for their much larger uncertainty. While their detection methods were proven robust with respect to known candidate MBCs, we note that their sample included no unambiguously cometary objects. Hence, it is difficult to compare our result to theirs, but the possibility of a steeply increasing number distribution for MBCs below the ∼1-km level is not ruled out. Indeed, two known candidate MBCs, 238P/Read and 259P/Garradd, have measured sub-kilometer diameters ([Hsieh et al. (2009b)]; [MacLennan and Hsieh (2012)]).

2.8.3 Future work: Photometric (absolute magnitude) variation as a function of orbital anomaly

As suggested by this article's title, extended-object analysis is only the first kind of cometary-detection method to which we intend on subjecting the PTF small-body data set. We hope future works will analyze the time- (and mean-anomaly-) varying absolute magnitude of small bodies over orbital-period baselines, which could potentially reveal even unresolved cometary activity.

Preliminary analysis of PTF photometry of main-belt comet P/2010 R2 (La Sagra), which include pre-discovery observations, shows a time-resolved ∼1.5-mag increase in absolute magnitude and a corresponding factor ∼5 increase in the dust-to-nucleus cross-section ratio, A_d/A_N. These results suggest PTF is capable of detecting intrinsic disk-integrated flux variation at the level of known candidate MBCs. Upcoming analyses of other known comets in our sample should confirm this robustness.

As shown in Figure 2.15, the orbital-coverage of PTF-observed known objects is far from complete. The orbital period of main-belt objects varies from about three to six years; a desirable prerequisite to orbital variation analysis is a comparable survey duration (especially to alleviate the bias against longer-period outer main-belt objects). The use of only two visible-band filters[7] gives PTF an advantage over other ongoing surveys[8], since conversion between wavelength bands introduces uncertainty when object colors are unknown. Thus, multi-filter data makes absolute magnitude comparison between epochs (already complicated by uncertainties in spin amplitudes and phase functions) even more difficult. Finally, a photometric variation analysis would benefit from the inclusion of null-detections, which are not currently a product of our kd-tree harvesting method, but should be implementable with a reasonable amount of modification.

[7]In fact mostly just one: 87% of the ∼2 million small-body observations in this work are in Mould-R-band, 13% in g'-band.

[8]To illustrate this point by comparison, of the ∼3 million small-body observations reported to the Minor Planet Center by *Pan-STARRS 1* (PS1) as of mid-2012, ∼40% are w-band (a wide-band filter covering most of the visible), while g-, r- and i-bands each represent ∼20% of the PS1 data.

Chapter 3

Small Near-Earth Asteroids

3.1 Introduction

A near-Earth asteroid (NEA) is by definition any asteroid with perihelion $q < 1.3$ AU and aphelion $Q > 0.983$ AU. From the largest NEA (of diameter $D \approx 30$ km) down to $D \approx 0.5$ km in size—for which the known population is largely complete—the cumulative size-frequency distribution (Figure 3.1) goes roughly as $N(D) \propto D^{-2}$, where $N(0.5 \text{ km}) \approx 10^4$. Harris ([Harris (2008)], [Harris (2013)]) presents these statistics, and describes how the original 'Spaceguard' goal to catalog 90% of all $D > 1$ km NEAs was achieved by the mid-2000s, while the current goal is to find 90% of all $D > 140$ m NEAs by 2020.

The incrementally-decreasing target size in the NEA census has been mostly motivated by risk mitigation. Over the quarter-century that began with our realization of an asteroid's role in the dinosaurs' extinction (e.g., [Alvarez et al. (1980)]) through to our fulfillment of the 1-km Spaceguard goal, the estimated risk of an individual's death from asteroid impact—initially believed comparable to that of a commercial airplane accident—dropped by an order of magnitude. Surveying to the currently-recommended $D > 140$ m can decrease this risk by yet another order of magnitude [Harris (2008)].

Hence, discovery of $D < 100$ m NEAs is likely irrelevant to any further reduction in the risk to human lives. However, the size-frequency distribution informing these estimates is uncertain across orders of magnitude in impactor size, and constrained on the

small end ($D \lesssim 10$ m) by infrasound detections of bolide fluxes ([Silber et al. (2009)]). Events like the Tunguska and Chelyabinsk airbursts [Brown et al. (2013)], which did not result in deaths but caused environmental or municipal damage, suggest impacts from 10–100 m objects qualify as 'natural disasters' that merit advance warning, and possibly prevention via space-based manipulation of hazardous NEAs.

Besides impact mitigation (e.g., [Ahrens and Harris (1992)]; [Lu and Love (2005)]), other space-based activities benefiting from small NEA discoveries include in-situ compositional studies [Mueller et al. (2011)] and resource utilization [Elvis (2014)]. NEAs have also been declared a major component of NASA's manned spaceflight program [Obama (2010)]. NEA rendezvous feasibility depends critically on mission duration and fuel requirements, these in turn are functions of the NEA's orbit and relative velocity (Δv) with respect to Earth ([Shoemaker & Helin (1978)]; [Elvis et al. (2011)]). Robotic missions may facilitate or complement the manned program, e.g., via the proposed Asteroid Retrieval Mission (ARM; [Brophy et al. (2012)]), though the selection criteria for ARM targets are highly-specific as well[1]. Natural temporary capture of meter-scale NEAs into Earth-centric orbits, if confirmed via the discovery of 'mini-moons' ([Grav et al. (2012))]; [Bolin et al. (2014)]), would present another appealing class of targets.

Cleary, discovery of 1–100 m sized NEAs is motivated by different (and more diverse) applications than those which have driven the census of larger NEAs. The discovery *method* often likewise differs. Most large NEAs were found via the 'tracklet' method of linking several serendipitously-observed positions within a night or across several nights. This is the basis of 'MOPS'-like detection software (*e.g.,* [Denneau et al. (2013)]), which in its present state is most efficient at detecting NEAs moving slower than $\sim$5 deg/day [Jedicke et al. (2013)]. Below this rate, an NEA's individual detections are nearly point-like for typical survey exposure times (e.g., 30–60s), and sufficiently localized on the sky given typical intra-night pointing cadences (e.g., 15–45 minutes). Hazardous NEAs occupy a range of orbits with moderate

[1]Generally, a suitable ARM target has $D \approx 7$ m, orbital semi-major axis 0.7 AU $< a <$ 1.45 AU, eccentricity $e < 0.2$ and inclination $i < 8°$, corresponding to a relative velocity $\Delta v \lesssim 6$ km/s [Jedicke et al. (2013)].

eccentricities, and so spend most of their time far from the Earth and Sun, where their sufficiently slow apparent motions allow them to be easily detected with this technique.

In contrast, the method of *streak detection* enables discovery of much smaller and closer (i.e., brighter and faster-moving) NEAs. Whereas slower-moving NEAs can be mistaken for main-belt asteroids, streaked objects are nearly 100% certain to be NEAs. Unlike the tracklet method, discovery via streak detection is possible on the basis of a *single* exposure via recognition of the streak morphology, meaning repeat visits to the same patch of sky are unnecessary and more area can be searched. Streak detection extends the NEA-discovery lifetime of a survey telescope whose aperture size renders it no longer competitive in the area of tracklet-based discovery. Lastly, streaked NEAs are typically 2 to 3 magnitudes brighter than those found by the tracklet method, making them far more convenient for non-sidereally-tracked follow-up from dedicated (including amateur-class) facilities.

Survey-scale application of the streak-detection method for NEA discovery was pioneered by [Helin and Shoemaker (1979)] using photographic plates on the Palomar 18-inch Schmidt telescope in the 1970s. [Rabinowitz (1991)] was the first to apply this method with CCD detectors in near real-time with the Spacewatch survey. Combining Spacewatch's streaked NEA detections (e.g., [Scotti et al. (1991)]) with its tracklet-detected NEAs [Jedicke (1995)] produced a debiased NEA number-size distribution [Rabinowitz et al. (2000)] spanning four orders of magnitude in size (10 km $> D >$ 1 m).

Figure 3.2 breaks down the number of streaked NEA discoveries as a function of time and survey, from 1991 through 2014-Oct. Here 'streaked' is taken to mean any detection wherein the length of the imaged streak is greater than 10 seeing widths. The counts in Figure 3.2 were compiled by first retrieving all NEA discovery observations from the Minor Planet Center (MPC) database and then using JPL's HORIZONS service [Giorgini et al. (1996)] to compute the on-sky motion at the discovery epoch. These rates were then converted into streak lengths in units of seeing widths, where the continental surveys all have assumed seeing $2''$ and Pan-STARRS

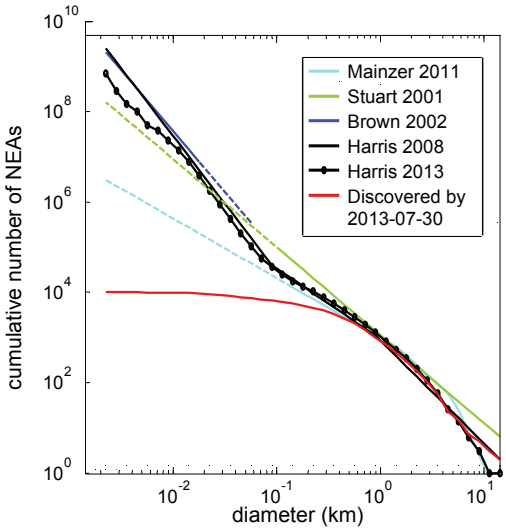


Figure 3.1: Cumulative NEA population distribution models compared to discovered objects. Plot adapted from a figure in [Ruprecht et al. (2014)].

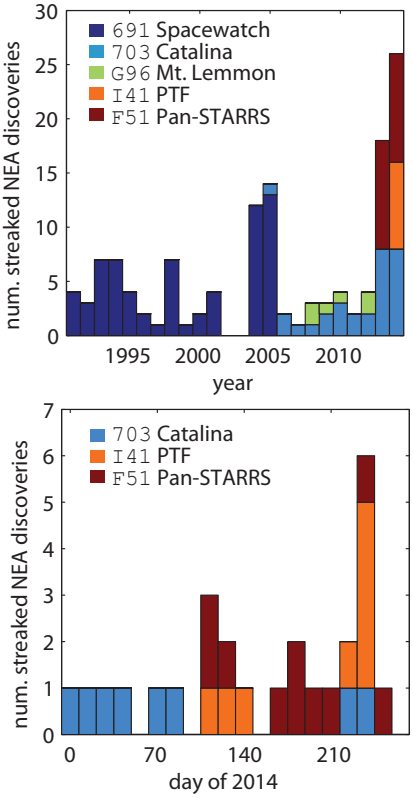


Figure 3.2: Number of streaked NEA discoveries as a function of time (bins include 1990–2014) and survey, where 'streaked' is here defined as any discovered streak greater than 10 seeing-widths in length (see text for details).

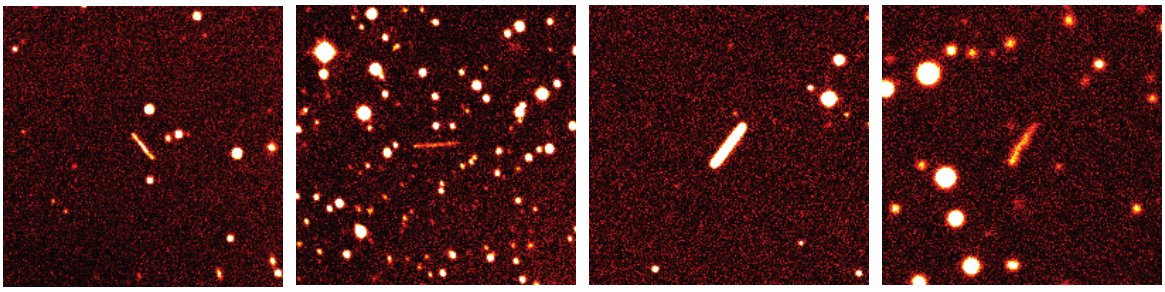


Figure 3.3: Some known small NEAs serendipitously detected by PTF. These observations were retrieved solely by computing these known objects' positions at the epochs of archival PTF images and visually verifying the streak's presence. All images are $200'' \times 200''$ with linear contrast scaling from -0.5σ to 7σ.

has assumed $1''$ seeing. The assumed exposure times come mostly from a table in [Larson (2007)], except for PTF and Pan-STARRS, which have assumed exposure times of 60s and 45s, respectively.

Before 2005, Spacewatch was the only contributor of significantly streaked NEA discoveries, and it is also the most prolific streaked-NEA discoverer overall. There are two likely reasons for this: (1) Spacewatch's relatively long 120s exposure time, and (2) the active role of a human screener ('observer') during data collection, as documented by [Rabinowitz (1991)]. The Catalina sky-survey also has a dedicated human operator to scan candidates and conduct same-night follow-up [Larson (2007)], which explains its similarly consistent contribution of streaked discoveries. Some major NEA surveys of the past two decades *not* contributing to the streaked discoveries in Figure 3.2 include LINEAR [Stokes et al. (2000)]—likely because of its short 8s exposures, as well as NEAT [Pravdo et al. (1999)] and LONEOS [Stokes et al. (2002)]—which to our knowledge lacked real-time human interaction with their respective data flows.

The years 2013 and 2014 marked a clear upturn in the discovery of streaked NEAs. The purpose of this chapeter is to document a new streak-discovery pipeline which has contributed in part to this increased discovery rate.

3.2 Overview of the PTF survey

3.2.1 Technical and operational characteristics

The *Palomar Transient Factory*[2] (PTF) is a synoptic survey designed primarily to discover extragalactic transients ([Law et al. (2009)]; [Rau et al. (2009)]). The PTF camera, mounted on Palomar Observatory's 1.2-m Oschin Schmidt Telescope, uses 11 CCDs (each 2K × 4K) to image 7.3 deg^2 of sky at a time (at 1.0″/pixel resolution). Most exposures (∼85%) use a Mould-R filter[3] (hereafter "R") with a 60-second integration time. Science operations began in March 2009, with a nominal 1- to 5-day cadence for supernova discovery and typical twice-per-night imaging of fields. Median seeing is 2″ with a limiting magnitude $R \approx 20.5$ (for 5σ point-source detections), while dark conditions routinely yield $R \approx 21.0$ [Law et al. (2010)]. About 15% of nights (near full moon) are devoted to an Hα-band imaging survey of the full Northern Sky.

In January 2013 the PTF project formally entered a second phase called the *intermediate PTF* ('iPTF'; [Kulkarni (2013)]). For most of this chapter we simply use 'PTF' to mean the entire survey, from 2009 to the time of writing (late 2014), though we note that PTF's NEA-discovery capabilities were conceived, funded, developed and commissioned entirely in this post-2012 'iPTF' period. This is partly because iPTF accommodates more varied 'sub-surveys' as opposed to a predominantly extragalactic program, including variable star and solar system science.

As will be detailed later (e.g., Figure 3.15), typical PTF pointings tend to avoid the ecliptic (and hence opposition) in accordance with its primarily non-solar system science objectives. In recent summer seasons, PTF has also spent the majority of its observing time imaging the dense galactic plane; many such galactic fields contain very high source densities and were not capable of being processed with the streak detection pipeline described below.

[2] http://ptf.caltech.edu

[3] The Mould-R filter is very similar to the SDSS-r filter; see [Ofek et al. (2012a)] for its transmission curve.

3.2.2 Previous solar system science with PTF

The present chapter discusses the first NEA-related (and first real-time) work with PTF solar system data; previous PTF solar system work analyzed archival observations of main-belt asteroids. [Polishook et al. (2012)] and later [Chang et al. (2014a)] used high-cadence data (which is uncommon in PTF) for 'pilot studies' of asteroid rotation lightcurves spanning consecutive nights. [Waszczak et al. (2013b)] mined PTF for all observations of known asteroids and then searched this data set for activity characteristic of 'main-belt comets' [Hsieh and Jewitt (2006)]. We used this database of known-object observations to extract detections of known streaking NEAs in PTF (Section 3.2.4). [Waszczak et al. (2013b)] also developed an original MOPS-like tracklet-finding routine which was later implemented in the real-time IPAC pipeline discussed below, but is otherwise unrelated to the streak-detection pipeline.

3.2.3 Real-time data reduction at IPAC

Since the survey's start, PTF has employed two separate data reduction pipelines serving distinct purposes. A real-time image-subtraction pipeline hosted at the National Energy Research Scientific Computing Center at Berkeley Lab (Nugent et al. in prep.) forms the basis of the extragalactic transient discovery program. A separate, archival-grade image-processing pipeline hosted at the Infrared Processing and Analysis Center (IPAC) at Caltech [Laher et al. (2014)] runs during the day and performs flat-fielding, bias-subtraction, source catalog generation, and astrometric and absolute-photometric calibration.

In early 2013, a real-time version of the IPAC image-processing pipeline was put into regular nightly operation. This initial version included daily automated batch submission of main-belt (and slow-moving near-Earth) asteroid observations to the MPC (both known objects and new discoveries). In addition to the above-mentioned image reduction features detailed by [Laher et al. (2014)], the real-time processing includes an original module for image subtraction [Masci (2013)], which uses a deep co-add of $\sim$20 previous PTF images that reaches $V \approx 22$ (i.e., a 'reference' or 'tem-

plate' image). The reference image is convolved with the new image's PSF kernel prior to subtraction, as described by [Masci (2013)]. The creation of this real-time IPAC pipeline precipitated the development of the streak-detection system discussed in this chapter.

3.2.4 Detections of known streaking NEAs

Early in the development of our streak detection system, we sought to extract all observations of known fast-moving NEAs from existing PTF data. We used the table of all predicted PTF sightings of all known asteroids compiled by [Waszczak et al. (2013b)], updated to include data through early 2014.

There are a total 539 predicted sightings (of 158 unique objects) for which the predicted motion was faster than $10''$/minute and the predicted magnitude was brighter than $V = 20$. For objects having predicted positional uncertainties greater than $10''$, the images were visually inspected around the predicted location for the presence of a streak. Because the $V < 20$ brightness criterion is based upon HORIZONS-predicted magnitudes, which have a typical accuracy of ~ 0.5 mag, in certain cases the actual magnitude was almost certainly fainter than $V = 20$. These particular predicted sightings (having good positional localization but are possibly too faint for detection) are still included as long as the predicted (point-source) magnitude is brighter than $V = 20$.

Figure 3.3 shows some examples of visually-confirmed PTF streak detections from this set of predicted sightings. Qualitative variations in morphology due to a differences in magnitude, streak length and seeing are apparent.

As described below, these 539 serendipitous sightings constituted the initial test bed for development of our streak detection algorithm. As of this writing (late 2014), PTF has acquired ∼90 new additional detections include unconfirmed PTF discoveries, confirmed PTF discoveries, and PTF-observed discoveries from other surveys in 2014.

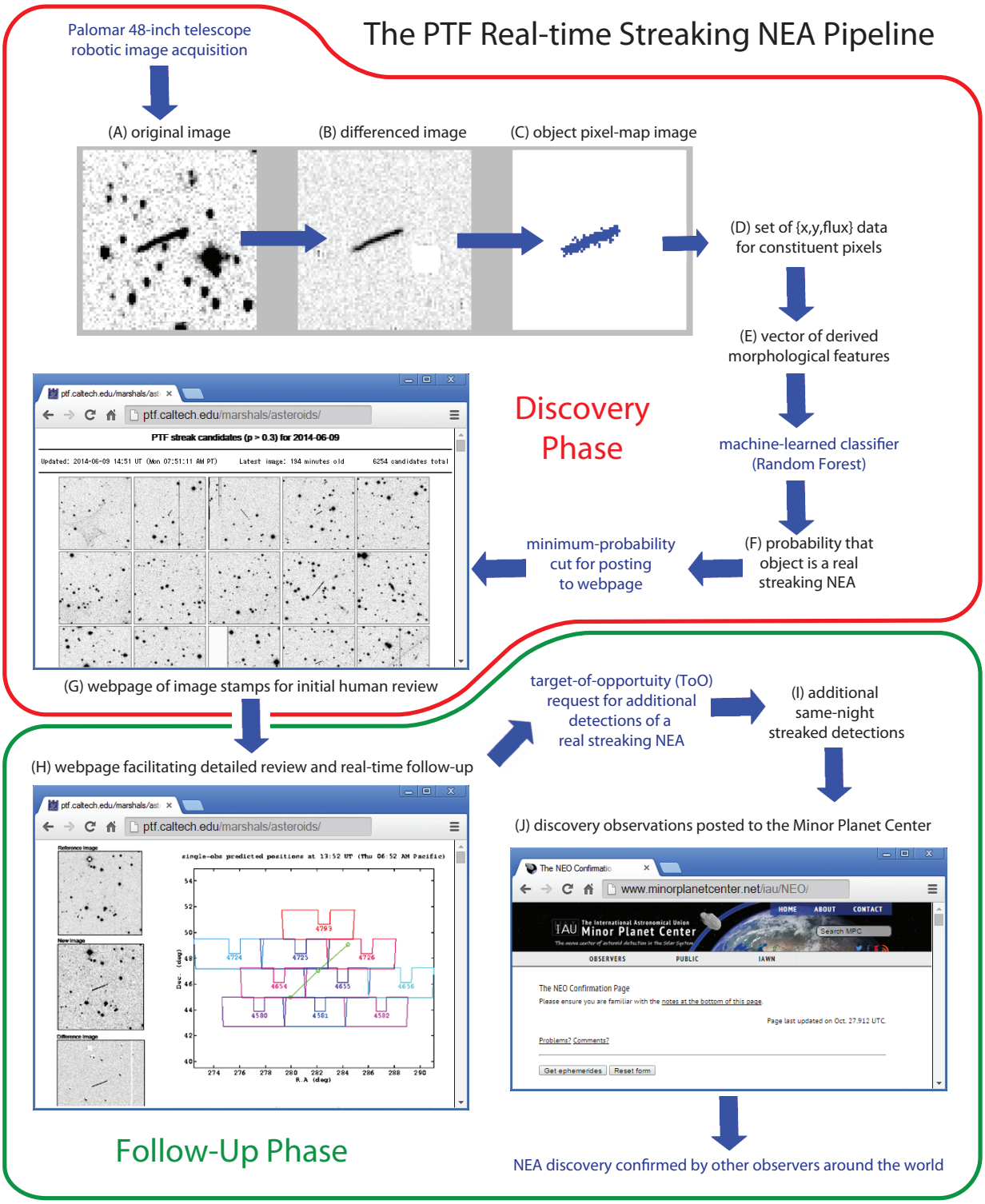


Figure 3.4: Flowchart depicting the PTF streaking-NEA discovery pipeline.

3.3 Streak-detection process

The principal steps of the streak-detection process are:

1. Image processing and subtraction of a reference image to produce a differenced image
2. Detection of candidate streaks as regions of contiguous pixels on the differenced image
3. Measurement of a set of morphological features describing each candidate streak
4. Filtering of likely non-real detections on the basis of their computed features
5. Human recognition of real streaks by reviewing images of the filtered candidates

The above five steps comprise the *discovery* phase and entail the creation of the data products labeled (A) through (G) in Figure 3.4. Upon discovery of a real streak, data products (H) through (J) are created as part of the *follow-up* phase, which we discuss later (Section 3.4).

Initial image processing and reference-image subtraction (first of the above-enumerated steps) are described by [Laher et al. (2014)] and [Masci (2013)], respectively. Step 2 involves identifying the pixels on the differenced-image belonging to candidate streaks. Whereas pixel-level data for point-source transients (e.g., supernovae or slow-moving asteroids) can be efficiently extracted with commonly used software such as *Source Extractor* [Bertin and Arnouts (1996)], streaked detections require a distinct approach as their image footprints contain many more pixels, often with much lower signal to noise per pixel. To meet this need we developed an original piece of software called `findStreaks`.

3.3.1 Object detection with `findStreaks`

3.3.1.1 Algorithm

The `findStreaks` software is derived from code originally created for the IPAC processing pipeline to identify and mask very long tracks in PTF exposures due to

satellites and aircraft [Laher et al. (2014)]. `findStreaks` was developed in the C programming language to maximize computing speed. The software first thresholds the image pixels above a local background noise level, then groups contiguous pixels into objects or 'blobs' (i.e., candidate streaks), and lastly computes morphological features for each object.

The differenced-image's local median background values are computed on a coarse grid with 64-pixel grid spacings and 129-pixel windows, with bilinear interpolation used to fill in the pixel values between the grid points. These median values are used to threshold the positive difference image at 1σ above the local median. All below-threshold pixels are discarded, and only above-threshold data are considered further (e.g., Figure 3.4 item C). Image-edge pixels are ignored (to avoid artifacts along CCD edges).

The `findStreaks` module arranges all contiguous blobs of pixels, each blob in one or more segments of computer memory, where adjacent pixels in the cardinal and diagonal positions are considered to be connected. For efficient memory management, the module is configured to handle up to 1 million memory segments, and up to 1000 pixels per segment. The sky-background-subtracted blob flux and instrumental magnitude are computed, along with their respective uncertainties. The median and dispersion of the pixel-blob intensity data are computed and subsequent morphological analysis is done only on pixels with intensities that are within $\pm 3\sigma$ of the median, where σ is given by half the difference between the 84th and 16th percentiles. A line is fit to all pixel positions in each blob, and the slope and y-intercept are obtained, as well as the linear correlation coefficient:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum_i (x_i - \bar{x})^2)(\sum_i (y_i - \bar{y})^2)}}. \tag{3.1}$$

Perpendicular distances from the linear model to constituent pixels are used to find the blob half-width, defined as half the difference between the 84th and 16th percentiles of these distances. With size and shape parameters now in hand, several hard filters are used to eliminate blobs that are not considered to be streaks. Blobs

Table 3.1: Morphological and other features saved for streak candidates.

feature	description
`pixels`	number of pixels associated with detected object
`length`	long axis length
`hwidth`	half-width
`dMax`	perp. distance of maximum-flux pixel from longest axis
`angle`	proper angle (in RA, Dec coords)
`median`	median pixel flux
`scale`	1σ variation in pixel flux
`slope`	slope (dy/dx in image coordinates) of fitted line
`correl`	correlation coefficient of fitted line
`flux`	total flux of object
`refDist`	distance from midpoint to nearest object in reference image
`refMag`	magnitude of nearest object in reference image
`epoch`	epoch (modified julian date)
`ra`	right ascension of object midpoint
`dec`	declination of object midpoint

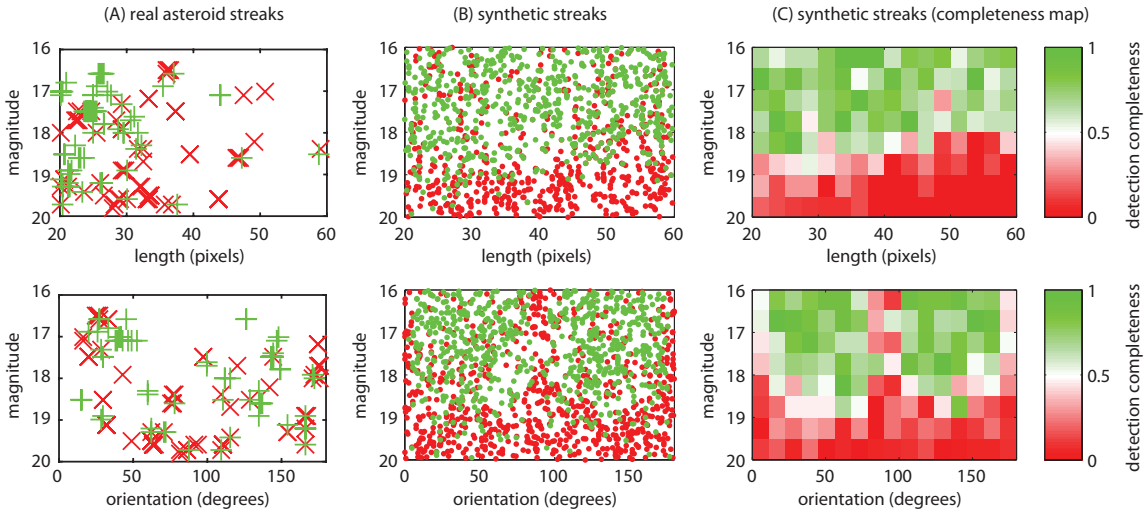


Figure 3.5: Successful detections (green) and failed detections (red) for both real asteroids and synthetic streaks. Here a 'successful detection' means an object was found by `findStreaks` at the predicted location having a measured length within four streak-widths of the predicted length. In the real data (leftmost plots), multiple detections of unique objects are often very close to one another in the 2D spaces plotted here, such that the total number of points discernible on the plot may appear less than actual.

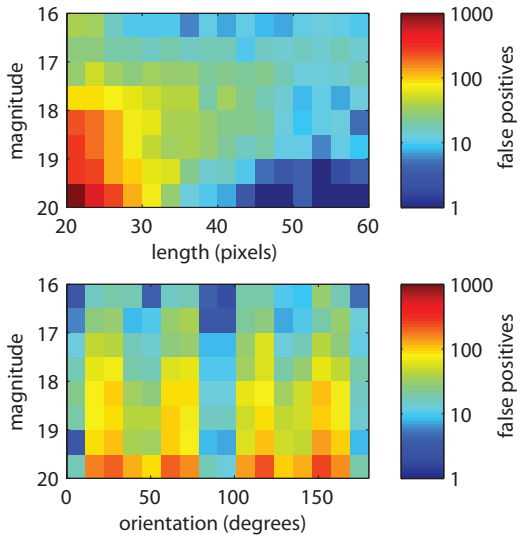


Figure 3.6: Distribution of false positive detections from `findStreaks`. The largest concentration of these 'bogus' detections are in the short and faint regime. Structure as a function of orientation angle (bottom) is due to a combination of the correlation sensitivity (see text) and pixel effects, wherein diagonal ($\pm 45°$-oriented) blobs are less likely to exist as their flux is diluted across more pixels.

containing more than 400,000 pixels, having long axes shorter than 9 pixels (i.e., 3.6 deg/day motion), or having half-widths larger than 16 pixels are discarded. Blobs for which the absolute value of the linear correlation coefficient is less than 0.5 are also discarded.

The `findStreaks` module outputs a table of streak metadata, where each table row corresponds to a streak detection. The real-time pipeline augments this table with additional columns including the proximity of the candidate to the nearest reference-image (stationary) object, as well as the brightness of this nearest reference-image object. Table 3.1 lists the 15 features currently retained for each candidate, and used in the classification stage that follows. This list of features will be updated to include additional morphological metrics in future versions of this software, but the results of this chapter only include analysis of the above-described 15 features.

3.3.1.2 Completeness and contamination

To ascertain `findStreaks`'s completeness and the number and nature of false positives it detects, we tested the software on a set of images containing both known real asteroid streaks (the 539 predicted sightings described in Section 3.2.4) and a large number of injected synthetic (simulated) streaks.

To generate each synthetic streak 'stamp', we first considered a 2D-Gaussian point-spread function of flux f, full-width at half-maximum (FWHM) θ, and center at (x_0, y_0):

$$\begin{aligned} &\mathrm{PSF}(x, y, x_0, y_0, f, \theta) = \\ &f \times \frac{4 \ln 2}{\pi \theta^2} \times \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{\theta^2} \times 4 \ln 2 \right) \end{aligned} \tag{3.2}$$

In terms of Eq. (1), a simulated asteroid streak of length L oriented at angle ϕ is given by

$$\begin{aligned} &\mathrm{Streak}(x, y, x_0, y_0, f, \theta, L, \phi) = \\ &\qquad \frac{1}{L} \int_{t=0}^{t=L} \mathrm{PSF}(x - t \cos\phi, y - t \sin\phi, x_0, y_0, f, \theta)\ dt \end{aligned} \tag{3.3}$$

[Vereš et al. (2012)] presents a similar streak model albeit with a slightly different analytical expression.

We evaluate the integral in Eq. (2) numerically over a grid with spacings $\Delta x = \Delta y = 0.05''$. Assuming the physical units of x and y are PTF-image pixels ($= 1.0''$), and assuming a typical PTF seeing value of $\theta \approx 2''$ (though we randomly vary θ along with other parameters, see below), the $0.05''$ grid spacing ensures the simulated streak is initially oversampled (by a factor of several tens) relative to the final (coarsened) image of the streak.

For each synthetic streak, the various model parameters in Eq. (2) are randomly drawn from flat distributions on the following intervals:

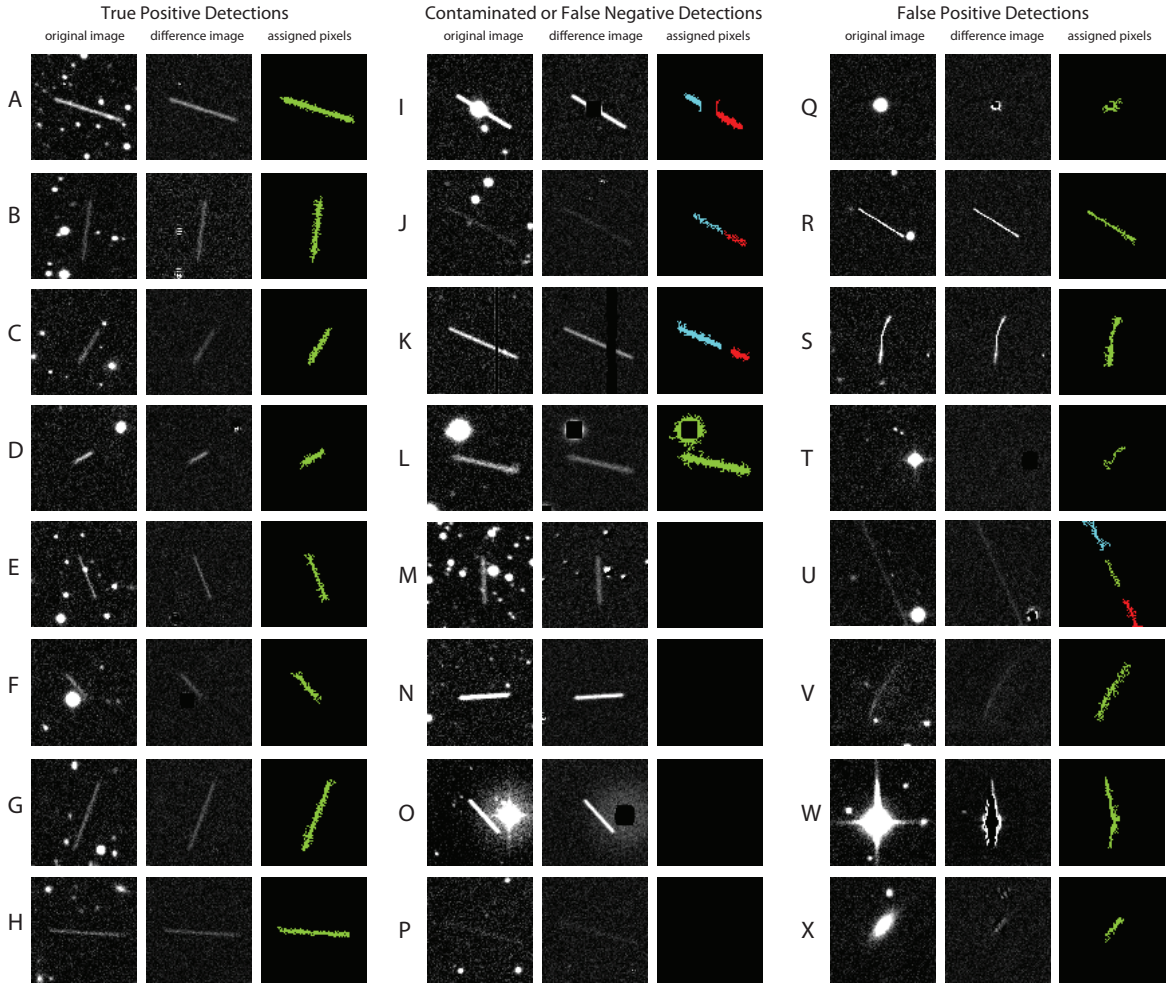


Figure 3.7: Examples of streak detections in PTF images. The third column, "assigned pixels", shows the pixels mapped to the object by `findStreaks`, wherein unique objects are distinctly colored. (I) Splitting due to saturated star (undefined pixels on difference image). (J) Splitting due to faintness. (K) Splitting due to bad column in difference image. (L) Extraneous pixels from nearby bright star halo. (M) Missed detection due to near-vertical orientation. (N) Missed detection due to near-horizontal orientation. (O) Missed detection due to large variation in background levels (star halo). (P) Missed detection due to faintness. (Q) Poorly-subtracted star false-positive. (R) Linear radiation hit. (S) Non-linear radiation hit. (T) False positive due to background noise. (U) Isolated segment of longer faint streak (e.g., due to a satellite). (V) Portion of optical ghost artifact. (W) Diffraction spike false positive. (X) Poorly-subtracted galaxy false-positive.

$$\begin{aligned} 0 < x_0 < 1 \qquad & 0 < y_0 < 1 \\ 1.4'' < \theta < 3'' \qquad & 10'' < L < 60'' \\ 0^\circ < \phi < 180^\circ \quad & 1800 \text{ counts} < f < 7200 \text{ counts}. \end{aligned} \tag{3.4}$$

A synthetic streak's flux f relates to its apparent magnitude m according to $m = m_0 - 2.5 \log_{10} f$, where m_0 is the zeropoint of the streak's host image. As the insertion of synthetic streaks into host PTF images is random (see below), it follows that the apparent magnitude m is not sampled from a *uniform* distribution, unlike the parameters f, θ, ϕ and L. The counts for f prescribed above roughly simulate 15 mag $< V <$ 21 mag for typical PTF zeropoints (given normal variations in sky background, extinction, etc.).

To coarsen each synthetic streak (prior to injection into an image), we evaluate the mean flux value in each $1'' \times 1''$ bin, equivalent to downsampling the initial simulated image by a factor 20. We round the counts in each resulting pixel to the nearest integer, and crop the streak image to a rectangular 'stamp' including all non-zero pixels. Lastly, to simulate shot noise, we replace the value of each non-zero pixel with a random integer sampled from a Poisson distribution whose mean is equal to the original pixel value.

We generated a set of 5,000 synthetic-streak image stamps following the above process, and then inserted these at random locations into the 539 PTF images containing each of the 539 predicted known streak sightings (Section 3.2.4). In particular, for each image the number of streaks injected was determined by drawing from a Poisson distribution with mean equal to 5. A set of that number of synthetic streaks was then randomly drawn (with replacement) from the pool of 5,000 and stamped into the image at randomly-chosen (x, y) coordinates. The total number of injected streaks was 2,631. We then processed each image with an offline version of the IPAC real-time image-differencing pipeline (Section 3.2.3) and ran `findStreaks` on the differenced images.

In the real-time streak detection pipeline, the output of `findStreaks` is subse-

quently subjected to machine-learned classification and human vetting. However, in the interest of initially assessing the completeness and reliability of `findStreaks` as an isolated module, we here simply (albeit arbitrarily) define a 'successful detection' (i.e. a true positive detection) as any case wherein `findStreaks` found an object whose measured center lies within a 15″ radius of the streak's true center, and the measured length minus the true length is less than four times the streak's measured width. The successful and failed detections according to these criteria are plotted in Figure 3.5.

Two trends evident from the synthetic streaks are limiting magnitude-vs.-length (Figure 3.5 top row) and lack of sensitivity to near-vertical or near-horizontal streaks (Figure 3.5 bottom row). In general the completeness drops sharply at a certain limiting magnitude; this limiting magnitude brightens from ∼19 mag at 20 pixels to ∼18 mag at 60 pixels. Streaks oriented very near to either 0° = 180° or 90° are much less reliably detected by `findStreaks` (at all magnitudes)—this is due the imposed hard limit on correlation ($|r| > 0.5$), a criterion which both near-vertical and near-horizontal streaks fail to satisfy.

The total number of candidate streaks returned by `findStreaks` in this test was 21,783, or an average of ∼40 per image, most being false positive detections (also referred to as 'bogus' detections later in this chapter). Figure 3.6 details the distribution of false positives in magnitude, length and orientation space. These plots indicate that the most common type of false-positive detections are faint and short, consistent with these contaminants being mostly star/galaxy subtraction artifacts and segments of extended, low-surface brightness objects like optical ghosts, space debris trails and bright-star halos. Figure 3.7 presents a gallery of examples of successful, failed, and contaminant detections.

Among the 539 predicted sightings of real streaking asteroids (see section 3.2.4) in the test images, a total of 240 were successfully detected by `findStreaks`. The left-side plots of Figure 3.5 show successful and failed detections in the same feature subspaces in which the synthetics are also plotted in Figure 3.5. A distinction between the y-axis-plotted 'magnitude' for the reals and that of the synthetics is that

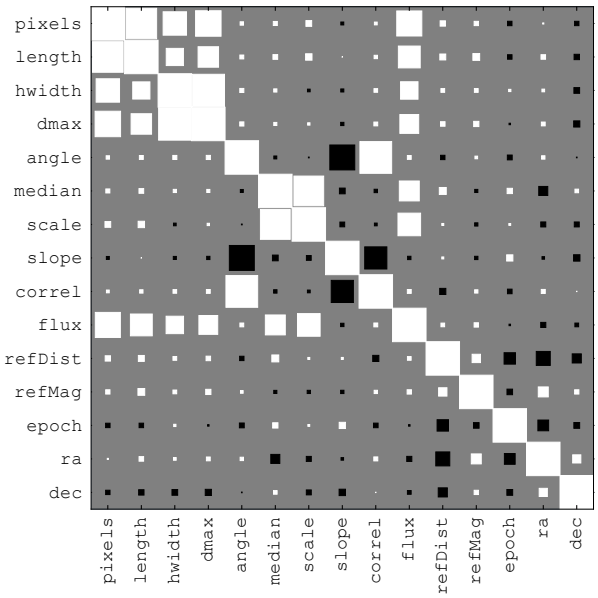


Figure 3.8: Correlation matrix for the 15 features (descriptions given in Table 3.1) used in the classification process. White squares indicate positive correlation, black indicate negative (anti-) correlation, and the area of each square indicates the magnitude of the correlation.

the magnitude of the reals is again the *predicted* brightness, accurate to $\sim$0.5 mag, whereas the synthetic magnitudes are more precisely known (even for the non-detected synthetics, as they still have a known flux and well-defined image zeropoint).

3.3.2 Machine-learned classification

3.3.2.1 Overview

As described above, a typical PTF image (single CCD) may contain several tens of false positive streak candidates, so that a full night of PTF observations—consisting of several thousand such images—may typically produce of order 10^5 raw candidate objects. This is far too many to screen manually by eye. Imposing simple filters on the measured morphological features (Table 3.1) can eliminate large subsets of false positives, but these hard cuts generally come at the cost of decreased completeness. A good example is the filter on the linear correlation coefficient condition ($|r| > 0.5$) discussed above, and the resulting insensitivity to near-vertical and near-horizontal streaks.

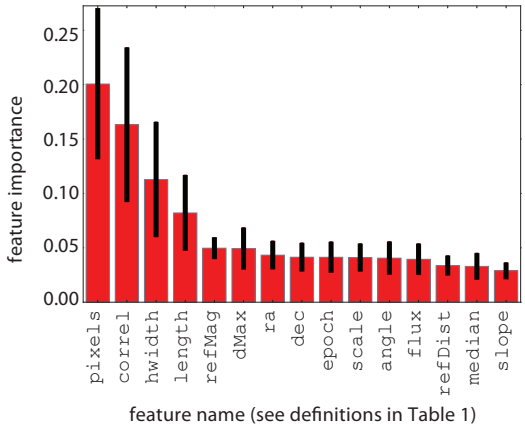


Figure 3.9: Importance of each of the 15 features (descriptions given in Table 3.1) used in the classification process. This number represents the fraction of training samples in which each feature contributes more by virtue of being at an earlier node splitting in the decision tree.

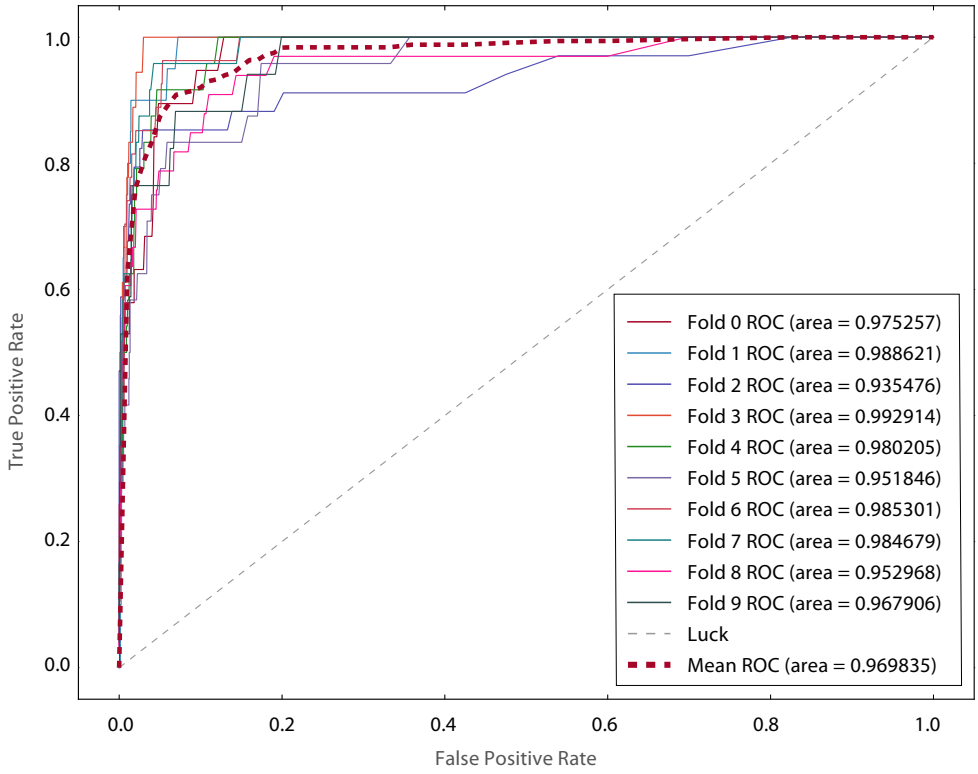


Figure 3.10: Classifier performance for each of the ten cross-validation trials. A plot in true-positive versus false-positive space is commonly referred to as a *receiver operating characteristic* (*ROC*) curve.

To address this issue we have trained and implemented a machine-learned classifier to discriminate real streaks from false positives. We adopt a supervised ensemble-method approach for classification, originally popularized by [Breiman et al. (1984)], specifically the *random forest* (RF) method [Breiman (2001)]. RF classification has extensive and diverse applications in many fields (e.g., economics, bioinformatics, sociology). Within astronomy in particular RF classification is one of the more widely-employed methods of machine-learning, though many alternatives exist. For example, [Masiero et al. (2014)] use the RF method for variable-star lightcurve classification, while others have approached this problem via the use of, e.g., support vector machines [Woźniak et al. (2004)], Kohonen self-organizing maps [Brett et al. (2004)], Bayesian networks and mixture-models [Mahabal et al. (2008)], principle component analysis [Deb & Singh (2009)], multivariate Bayesian and Gaussian mixture models [Blomme et al. (2011)], and thick-pen transform methods [Park et al. (2013)].

For general descriptions of RF training and classification, we refer the reader to [Breiman (2001)], [Breiman & Cutler (2004)], and the many references cited by [Masiero et al. (2014)]. Our use of a RF classifier is particularly motivated by its already-proven application to the discovery and classification of astrophysical transients in the same PTF survey data [Bloom et al. (2012)].

Streak candidates in PTF images are cast into a vector of quantitative morphological and contextual *features*, namely the 15 features listed in Table 3.1. Given a large set of such candidates, these metrics define a multi-dimensional space, which can be hierarchically divided into subspaces called *nodes*. The smallest node—also known as a *leaf*—is simply an individual candidate. Given a set of leaves with class labels, i.e., a *training set*—one can build an ensemble of trees (called a *forest*), each tree representing a different, randomly-generated partitioning of the feature space with respect to a subset of the total training sample (and a subset of the total list of features). The forest allows one to assign a probability that a given vector of features belongs to a given class. For the PTF candidates, were are interested in a binary classification, i.e., whether the candidate is real or 'bogus'. [Bloom et al. (2012)] coined the term `realBogus` to describe this binary classification probability. In the present work we

are essentially adapting Bloom et al.'s `realBogus` concept to the problem of streaking asteroid discovery.

3.3.2.2 Implementation and training

We employ a Python-based Random Forest classifier included as a part of the `scikit-learn` Python package[4]. Specifically, we use the `ExtraTreesClassifier` class in the `sklearn.ensemble` module. This particular code is an implementation of the 'extremely randomized trees' method [Geurts et al. (2006)], a variant of the Random Forest method containing an added layer of randomness in the way node-splitting is performed. Specifically, ExtraTrees chooses thresholds randomly for each feature and picks the best of those as the splitting rule, as opposed to the standard RF which picks thresholds that appear most discriminative. The additional randomization tends to improve generalization over the standard RF algorithm, this was verified empirically for our streak data.

Our training data consists of all candidate streak detections from the 539-image synthetic-injection test described in Section 3.3.1.2. This includes 240 real detections (out of the 539 predicted sightings from Section 3.2.4), 1,285 synthetic detections (out of the 2,631 total injected) and 20,072 bogus detections. Various examples of these bogus (false-positive) detections are shown in the right column of Figure 3.7, while their distributions in magnitude-vs.-length and magnitude-vs.-orientation space are shown in Figure 3.6.

Among the 15 features (Table 3.1) describing the streak candidates, several of the features exhibit some level of correlation, as shown in Figure 3.8. Most correlations are reasonable as they express the relationship between geometrically-similar quantities: the length of a streak is generally correlated with the number of pixels, and the fitted linear slope correlates with the proper angle. A strong correlation between `median` and `scale` (measures of flux signal and noise, respectively) is simply a expression of the Poisson noise associated with photon counting. Assessing the correlation between features aids in the interpretation of relative feature importances (Figure 3.9) derived

[4]http://scikit-learn.org

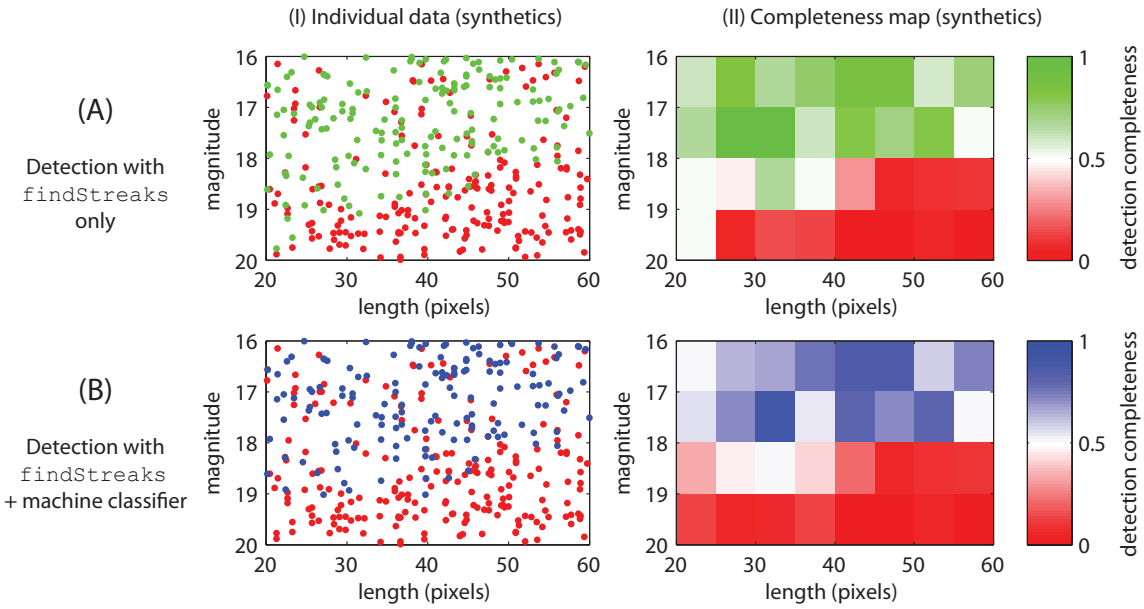


Figure 3.11: In the top row (`findStreaks` only), detection is again defined as the presence of an object whose length is within four streak widths of the true length, as in Section 3.3.2.1. In the bottom row (`findStreaks` plus the classifier), detection is defined as the presence of an object of length within four streak widths of the true length *and* a classification score of $p > 0.4$.

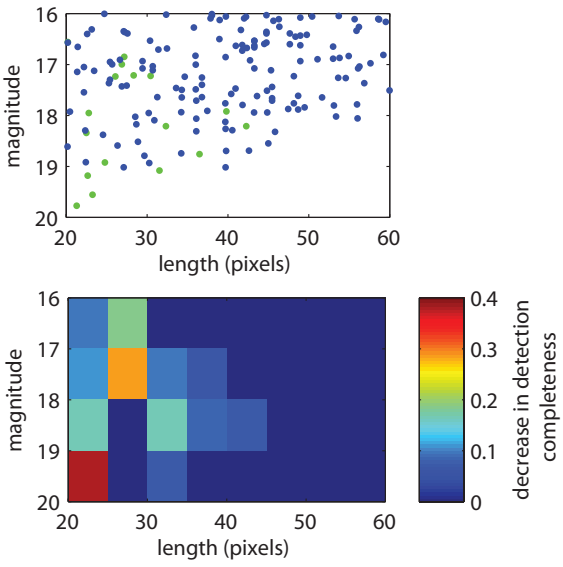


Figure 3.12: Loss in detection completeness due to the machine classifier—i.e., like Figure 3.10 except considering only those candidates that were first positively detected by `findStreaks`).

during the training process (described below). In particular, among the top four most discriminative features (according to Figure 3.9), three are significantly correlated (`pixels`, `hwidth` and `length`).

The classifier training consists of a 10-fold cross-validation (i.e., bootstrapping) process, wherein we split the data (reals, synthetics and boguses) into 10 disjoint sets using stratified random sampling. Then, in each cross-validation fold, we train using 9 of the sets and test on the remaining one—however, we exclude the synthetics from this test sample. In each of the ten cross-validation trials, the classifier outputs a classification probability for each object in the test sample, and we track the true positive rate (TPR; fraction of real streaks accurately classified as reals) as a function of the false-positive rate (FPR; fraction of bogus streaks inaccurately classified as reals). In astrostatisics TPR is also commonly called *completeness* while FPR is equivalently one minus the *reliability*. The results of the separate trials, as well as the averaged result, are shown in Figure 3.10. By tuning the minimum classification probability (i.e., the `realBogus` score) used to threshold the classifier's output, one effectively moves along the hyperbola-shaped locus of points in TPR-vs.-FPR space seen in the plot.

Several parameters can be adjusted or tuned when working with a random forest classifier. First is the number of decision trees generated during the learning stage. Classification accuracy typically increases with the number of trees and eventually plateaus. Most applications employ hundreds to thousands of trees; here we found that 300 trees provide sufficient performance. Another tunable parameter is the number of randomly-selected features (out of the 15 total here considered) with respect to which nodes are split in building the decision trees. [Breiman (2001)] recommends using the square root of the number of features; however, here we found optimal accuracy when splitting with respect to *all* 15 features. Other parameters that can be tweaked are the maximum depth of a tree, the minimum number of samples per leaf, the minimum number of samples used in a split, and the maximum number of leaf nodes. We do not constrain any of these parameters, meaning we allow: trees of any depth, with any number of leaf nodes, leaf nodes consisting of a single sample, and

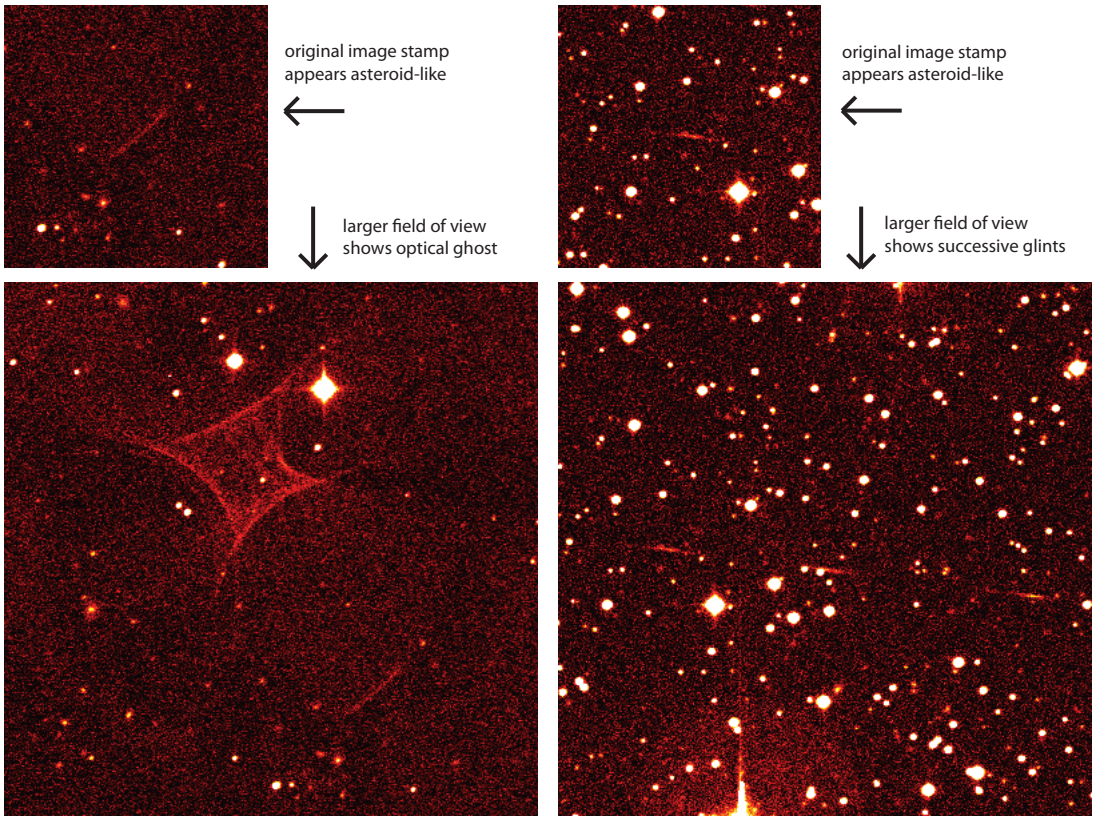


Figure 3.13: Example false positive detections in which the original $200'' \times 200''$ image stamp looks like a real asteroid streak, but the larger field of view clearly indicates the nature of the bogus detection. *Left*: Filament of an optical ghost. *Right*: Glint segment, e.g., from a fast-moving rapidly-rotating piece of space debris. If additional candidates from these larger false-positive objects also appear on the scanning page, their common exposure timestamp implies their stamps will appear adjacent to one another, facilitating their identification as bogus detections.

splits based on the minimum of 2 samples.

3.3.2.3 Post-training performance

In addition to tracking the classifier's performance during the training cross-validation trials, after training we subjected the classifier to a new sample of ∼400 synthetics. These newly-generated synthetics were injected into the same 539 test images using the same procedure described in section 3.3.1.2. Given the distinct random numbers used in this run, these synthetics are distinct from those that were used in training, and appear at different locations on the PTF images.

As was done in cross-validation, the purpose of this post-training trial was to ascertain the detection completeness, though this time using synthetics (which were used previously for training but not testing). Another difference is that we now consider completeness for a fixed classification probability threshold ($p > 0.4$) and do so as a function of magnitude and length (similar to the analysis done for `findStreaks` in Section 3.3.1.2).

The top plots of Figure 3.11 show the same information as was shown in Figure 3.5, albeit for this new sample of synthetics (and at slightly coarser resolution). Namely, we first examine the completeness delivered by `findStreaks` alone, and again see the limiting magnitude versus length trend. In the bottom plots of Figure 3.11, we show detection completeness for the same sample only this time for the combined `findStreaks` plus machine classifier system. In other words, all the blue data points in the lower left Figure 3.11 plot were both successfully detected by `findStreaks` *and* were subsequently classified as real with a probability $p > 0.4$.

In Figure 3.12, we again show data from the same synthetics sample, this time plotting the loss in absolute detection completeness due solely to the application of the machine classifier. In the top plot of Figure 3.12, green data points were successfully detected by `findStreaks` but did not score high enough ($p > 0.4$) in the classification stage. The 2D histogram below it shows that the most significant loss in completeness occurs for short faint streaks. Likely not coincidentally, this region suffers from the largest number of bogus `findStreaks` detections, as indicated by Figure 3.6. Integrating over all bins in this magnitude-vs.-length histogram, we observe an average completeness drop of $\sim$0.15, consistent with Figure 3.10 for a true positive rate of $\sim$85% accompanying a false-positive rate of $\sim$5%.

3.3.3 Web-based screening interface

The final component of the discovery portion of the PTF streak-detection pipeline consists of a webpage for human vetting of image stamps of streak candidates to which the classifier has assigned a high probability of being real. Figure 3.4 includes

a screenshot of this webpage. Given the $\sim$5% false-positive rate quoted in the preceding paragraph and the $\sim 10^5$ detected candidates accumulated in a typical night (cf. Section 3.3.2.1), this webpage displays on average several thousand candidates per night.

Including operations on Palomar Mountain, the data transfer from Palomar to Caltech, and the IPAC real-time processing pipeline (Section 3.2.3) a typical lag-time of $\sim$30 minutes (approx. $\pm$10 minutes) elapses between the acquisition of exposures with the PTF camera and the posting of streak candidates from said exposure to the scanning webpage. The image stamps have fields of view of $200'' \times 200''$ with linear contrast scaling from -0.5σ to 7σ (as in Figure 3.3). Undifferenced images are reviewed as opposed to the differenced images, to better provide context to the scanner and enable him/her to visually assess the observing conditions (i.e. the density and image quality of background stars).

The kinds of false positives commonly encountered on the scanning webpage include all of those shown on the right-hand side of Figure 3.7. Image stamps are viewed in chronological order, so that candidates from a common image appear consecutively on the scanning page. This enables rapid recognition of false positives of a common origin. For example, multiple segments of a long satellite trail, large optical ghost, or artifacts from a poorly-subtracted or high stellar density image will appear together and are thus easily dismissed. Artifacts that do not appear in groups, such as cosmic ray hits, background sky noise and poorly-subtracted galaxies, are rapidly visually dismissed as well. A full night's set of candidates (several thousand) can be reliably reviewed by a trained scanner in 5–10 minutes, though the reviewing time is distributed over the during observing session, as the webpage is refreshed every 20–30 minutes.

Clicking on the image stamp of a candidate streak presents another webpage with more detailed information including astrometry, photometry, `realBogus` score, image stamps of the differenced and reference images, and a larger field of view around the detection. Certain types of false positives are more easily identified using this additional information, including portions of optical ghosts and periodically-glinting

Table 3.2: PTF discoveries of streaked NEAs (between 2014-May-01 and 2014-Dec-01)

name	date	# PTF detections	diameter (m) (7% albedo)	inclination (deg)	eccentricity	semi-major axis (AU)	min. Earth orbit intersect. (LD)	Δv w.r.t. Earth (km/s)	speed ($''$/min)	V magnitude
2014 WS_7	Nov-19	3	17.4	8.687	0.47655	1.8918	4.36	6.1	25.7	18.7
2014 WK_7	Nov-18	4	166	23.91	0.36071	1.5605	3.38	8.4	39.8	16.7
2014 UL_{191}	Oct-30	2	66.2	2.070	0.65017	1.7553	3.73	7.8	66.9	17.1
2014 ST_{223}	Sep-23	4	15.2	5.875	0.17045	1.0532	1.56	5.2	31.8	18.4
2014 SC_{145}	Sep-23	3	36.4	20.17	0.20651	1.2216	6.61	7.8	28.4	19.0
2014 SE	Sep-16	2	43.4	20.02	0.17187	1.2428	13.5	7.9	24.7	18.7
2014 RJ	Sep-02	2	41.8	19.57	0.27083	1.4086	7.39	7.4	37.2	17.9
2014 LL_{26}	Jun-09	2	43.7	9.182	0.10177	1.1427	5.25	5.4	29.1	17.2
2014 KD	May-17	4	66.2	5.238	0.54605	2.1519	7.74	6.2	30.0	17.5
2014 JG_{55}	May-10	4	7.26	8.739	0.41257	1.5843	0.336	5.8	60.0	18.1

space debris. This summary page also contains information for real-time follow-up, as discussed in the next section.

3.4 Follow-up and reporting of discoveries

Once a real streak is discovered in PTF via the steps outlined in the previous section, we trigger real-time follow-up with the same telescope. Its wide field of view (Section 3.2.1) makes the PTF camera particularly well-suited for recovering fast-moving NEAs within a few hours of an initial detection. As described below, the follow-up process effectively interrupts the nominal robotic survey by injecting high-priority exposure requests into the queue. The final step involves reporting observations to the Minor Planet Center to facilitate subsequent confirmation and follow-up worldwide.

3.4.1 Target-of-opportunity (ToO) requests

The sequence of PTF fields observed on any given night is determined in real-time by a robotic scheduler: the P48 Observatory Control System (OCS) described by [Law et al. (2009)]. The robot takes as input a list of fields, generally prescribed by a human operator per lunation, and attempts to optimize exposure conditions (distance from moon, airmass, etc.) while also maintaining a specific cadence—predominantly two or three exposures per field per night separated by $\sim$40 minutes (optimal for supernova discovery). As noted in Section 3.2.1, in recent years (during the iPTF phase), fields and cadences have often been allocated to distinct experiments, though all exposures still adhere to a fixed tiling of fields, with 60-second integrations in either R- or g-band.

All PTF exposures are processed by the streak detection pipeline if they have a reference image available (required for image differencing, see Section 3.2.3). Upon recognition of a single detection of a likely real NEA streak on the scanning webpage, the human reviewer immediately checks the webpage for additional serendipitous detections in other PTF exposures acquired that night. If a second detection of the

same streak is found, the observations are immediately sent to the MPC's Near-Earth Object Confirmation Page (NEOCP)[5].

Lacking a second detection, the reviewer uses tools integrated into the scanning webpage to trigger *target-of-opportunity* (*ToO*) exposures to secure additional detections. Figure 3.4 shows a screenshot of the webpage's streak position estimation tool, which uses a linear (great-circle) extrapolation assuming motion in either direction, overlaid on the PTF tile grid. A PHP script redraws the plot to the current time when refreshed by the user.

Once a list of fields potentially containing the streak has been identified (typically between one and a few fields), a text-based email sent to the telescope robot inserts the fields into the queue with very high weight. This email may additionally prescribe repeat exposures of the fields with some specified cadence, filter, or maximum airmass. The ToO exposures typically are acquired within 5–10 minutes of the request, depending on factors such as slew time and the need to change filters. The email-based ToO-system for PTF was originally designed for (and proven on) the discovery of optical afterglows of gamma-ray bursts [Singer et al. (2013)].

Apart from having been manually triggered, the ToO exposures are otherwise identical to routine PTF survey images in that they are sidereally tracked, 60-second R- or g-band images aligned to a fixed tile grid of the sky (as opposed to, e.g., being centered on the NEA's predicted position). Having acquired the ToO exposures, any additional detections of the streak are automatically extracted with the same streak-detection pipeline and will appear on the scanning webpage along with the rest of the night's candidates. Observations are sent to the MPC once two or more detections have been secured.

3.4.2 Initial NEA discoveries

The full streak-discovery system, incorporating the IPAC real-time data products, `findStreaks` and the trained machine classifier, began real-time operations 2014-

[5]http://www.minorplanetcenter.net/iau/NEO/toconfirm_tabular.html

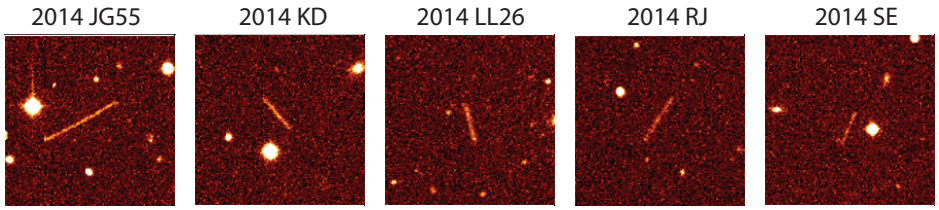


Figure 3.14: Discovery images of the first five streaked NEAs found by PTF.

Table 3.3: iPTF sub-surveys containing streaked-NEA discovery exposures.

name	sub-survey in which NEA was discovered	filter	degrees from opposition
2014 WS_7	Permanent Local Galaxies	R	28
2014 WK_7	TILU K2 Campaign	g	73
2014 UL_{191}	TILU Fall 2014	g	47
2014 ST_{223}	Opposition NEA search	g	18
2014 SC_{145}	RR Lyrae	R	32
2014 SE	RR Lyrae	R	21
2014 RJ	TILU Fall 2014	g	35
2014 LL_{26}	Star-forming low-cadence	R	9
2014 KD	TILU Spring 2014	R	49
2014 JG_{55}	iPTF14yb follow-up	R	34

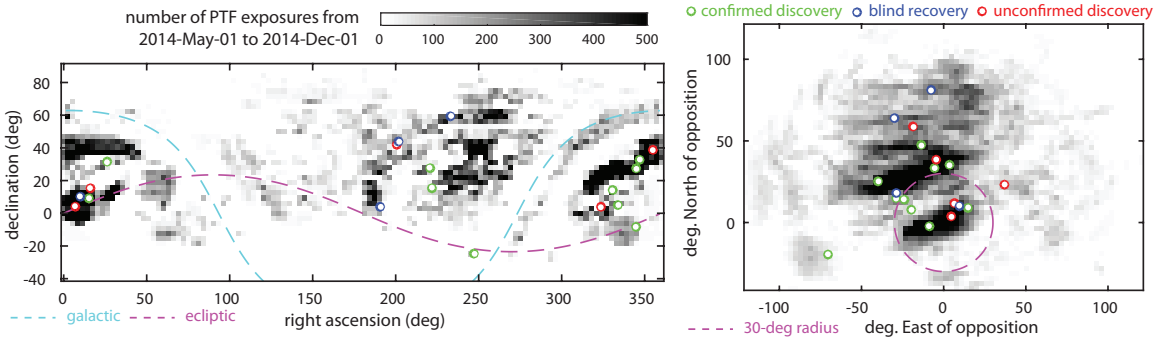


Figure 3.15: Distribution of PTF exposures (*left*: in sky coordinates, *right*: with respect to opposition) and streaked NEA detections (*right*: with respect to opposition) from 2014-May-01 through 2014-Dec-01. The grayscale scalebar maps the density of PTF exposures in both plots. Exposures for which realtime streak-detection was not performed are not included (e.g., fields lacking reference images or with too high source density on the galactic plane).

May-01. About a week later, the first PTF streaking NEA discovery was made (2014 JG_{55}). Passing at one-third of a lunar distance, this object is also the smallest and closest-approaching NEA yet discovered by PTF.

The largest streaking NEA discovered by PTF to date is 2014 WK_7, at $H = 22.4$ mag ($D \approx 166$ m), while the PTF discovery having an orbit most suitable for ARM (see the [Jedicke et al. (2013)] criteria in Footnote 1) is 2014 ST_{223}, though this object is probably too large for ARM.

Table 3.2 details the ten total streaking-NEA discoveries made by PTF as of 2014-Dec-01. Nearly all of these (the one exception being 2014 LL_{26}) were followed up and confirmed by multiple observatories within 24 hours. A total of 25 different observatories have provided follow-up observations within 24 hours of at least one of the NEA discoveries listed in Table 3.2. After the sun has risen in California, most short-term follow-up of PTF discoveries occurs from Japan and Europe (occasionally Australia), as most longitudes west of Palomar fall in the Pacific.

Figure 3.15 shows the discovery position of the ten NEAs in Table 3.1 relative to opposition. Most of the objects were found within 40° of opposition. An outlier is 2014 WK_7, which was discovered 73° from opposition (phase angle 71°), though this NEA is also an outlier in the sample in terms of its size.

Table 3.3 lists the various sub-surveys (also known as 'iPTF experiments', see Section 3.2.1) to which the NEA discovery exposures belong. Here 'TILU' stands for Transients in the Local Universe'. A key point here is that nearly all of PTF's streaked NEA discoveries to date have been made in images originally purposed for non-solar system science. A dedicated iPTF experiment designed to maximize the area covered around opposition was carried out for several nights in Fall 2014, though only one exposure from said program produced a discovery (2014 ST_{223}, which, as mentioned above, is the most ARM-like PTF discovery to date).

All follow-up was unsolicited apart from having posted the discoveries on the NEOCP, and attests to the dedication of the worldwide NEA follow-up community. We note however that, while they are on the NEOCP, PTF-discovered streaking NEAs are consistently the *brightest* on the list—all were $V \leq 19$ mag—whereas most of the

50+ objects typically found on the NEOCP have $V \geq 20$ mag. It is therefore not surprising that more follow-up facilities are able and willing to recover these bright objects as compared to the typical faint and slow NEOCP candidates.

3.4.3 Blind real-time recovery of known NEAs

There are several options for querying a given R.A., Dec., and time to search for a match (within some radius) to an asteroid with a known orbit; these include MPChecker[6], JPL's HORIZONS[7], and PyMPChecker [Klein et al. (2009)]. However, those scanning the PTF streak candidates in real-time are discouraged from checking if a detected streak is a known object prior to obtaining ToO follow-up and submitting the observations to the NEOCP. One reason is that the above mentioned query tools are not necessarily reliable for fast-moving objects, and will not always return a match even if the object has a well-determined orbit. Another reason is that the ToO-submitting procedure, while simple and straightforward, requires efficiency and efficacy on the part of the scanner and so should be practiced as often as possible. Lastly, the MPC encourages submission of unidentified known objects as it allows them to directly assess our program's detection capabilities (e.g., our astrometric accuracy).

As of 2014-Dec-01, a total of four previously-discovered NEAs have been blindly detected by PTF as streaks and submitted to the NEOCP: 2014 HL_{129} (May-02); 2010 JO_{33} (May-08); 2014 WF_{108} (May-27; to date the only 'potentially hazardous asteroid' blindly detected as a streak by PTF in real-time); and 2014 SE_{145} (Sep-23).

3.4.4 Unconfirmed discoveries

A total of five PTF objects posted to the NEOCP (between May-01 and Dec-01) did *not* receive external follow-up, meaning they never obtained confident orbit solutions and thus were not assigned provisional designations by the MPC (Table 3.4 and

[6] http://www.minorplanetcenter.net/cgi-bin/checkmp.cgi
[7] http://ssd.jpl.nasa.gov/sbfind.cgi

Table 3.4: Unconfirmed PTF streak discoveries (from 2014-May-01 to 2014-Dec-01)

NEOCP name	date found	num. obs.	speed ($''$/min)	V (mag)	notes
PTF5i5	May-04	2	46.8	19.5	
PTF9i2	Jul-08	2	36.9	17.9	85% moon, near dawn
PTF3k8	Sep-23	2	64.9	18.0	likely satellite
PTF8k2	Sep-25	3	27.6	18.9	
PTF7l3	Oct-25	2	30.1	17.5	

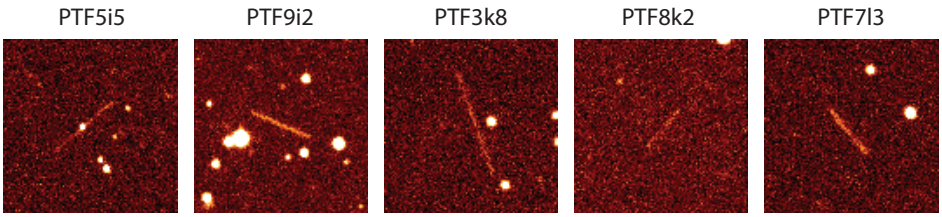


Figure 3.16: PTF streak discoveries that were posted to the NEOCP but never received external follow-up.

Figure 3.16). For four of these unconfirmed objects, PTF had submitted only two observations to the NEOCP. We note that 4 out of the 10 *confirmed* objects (Table 3.2) also were reported with only two observations, from which we naively conclude that a two-observation discovery has only a 50% probability of being successfully followed-up (for three- and four-observation discoveries the recovered fraction increases to 66% and 100%, respectively).

While in reality the recovery probability depends also on the temporal spacing of the observations, the object's speed and magnitude, and the availability of follow-up resources (e.g., less facilities operate around full-moon), the number of observations alone seems to be a useful indicator of the recovery likelihood. Users of the PTF real-time scanning and ToO system attempt to obtain at least three observations for discoveries, though this is not always possible, e.g., for discoveries made early in the night in the western sky, or just before dawn. Occasional technical issues with the real-time processing and/or ToO system also can hinder PTF self-follow-up.

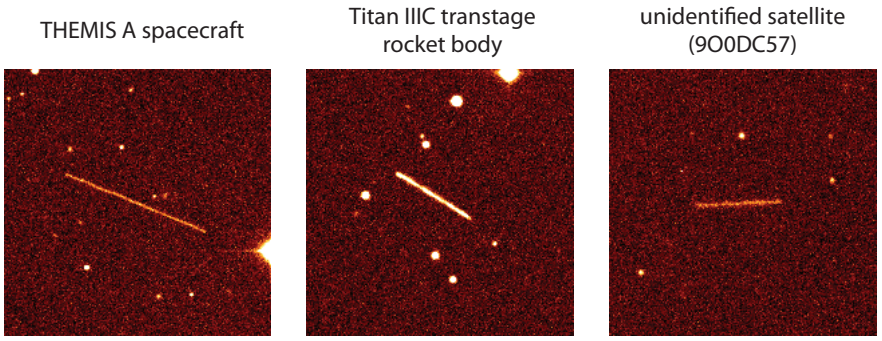


Figure 3.17: Artificial satellites detected as streaks by PTF (identifications provided by the MPC).

3.4.5 Artificial satellites

Many distant Earth-orbiting artificial satellites can, at certain parts of their orbit, appear consistent with an Earth-approaching NEA. Our streak-recognition pipeline has on several occasions detected such satellites. Figure 3.17 shows some examples, including one of the THEMIS mission spacecraft studying the Earth's magnetosphere [Angelopoulos (2008)] and a Titan IIIC rocket body. The MPC's automated observation-ingestion processes outputs known artificial satellite matches to NEOCP submissions (as was the case for the three in Figure 3.17), though in some cases the object will be posted to the NEOCP and remain on the list for some time prior to its recognition as artificial. Three examples of the latter were PTF7i2, PTF8i6, and PTF0n2.

While we see the same value in blind reporting of artificial satellites as we do blind reporting of known NEAs (Section 3.4.3), some high-orbit satellites have geosynchronous orbits and can therefore appear in the same area of sky for many consecutive nights. An example is the THEMIS spacecraft, whose apogee was coincident with opposition, causing it to be repeatedly observed by PTF in autumn 2014. For routine identification of known satellites, we have therefore adopted the useful software tool `sat_id` by Project Pluto[8].

[8]http://www.projectpluto.com/sat_id.htm

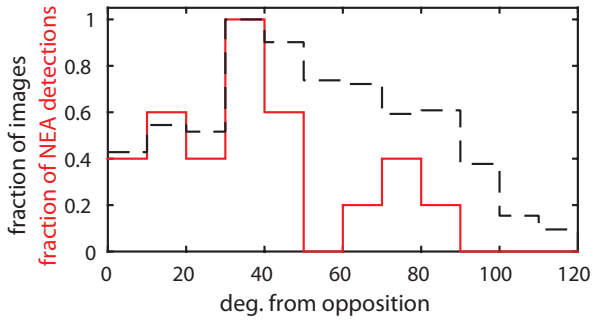


Figure 3.18: Normalized distributions of PTF images and streaked NEA detections with respect to opposition. The 19 NEAs included here consist of new discoveries, blind recoveries and the five unconfirmed discoveries. See Figure 3.15 for the two-dimensional distribution.

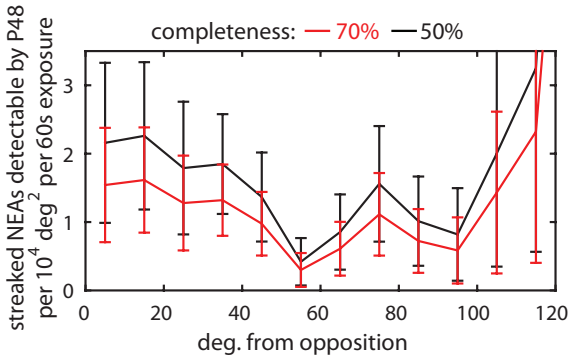


Figure 3.19: Estimates of the number of streaked NEAs detectable by P48 as a function of distance from opposition. Computed using the data in Figure 3.18 and Equation (3.5).

3.5 De-biased detection rate

The right panel of Figure 3.15 shows the distribution of streaked NEA detections (including confirmed and unconfirmed discoveries as well blind recoveries). In this section we use this sample of detections and the distribution of PTF exposures with respect to opposition to derive the de-biased streaked-NEA detection rate as a function of radial distance from opposition. Figure 3.18 shows the same data as in Figure 3.15, removing the azimuthal information to only show the one-dimensional radial distributions.

We seek to estimate the frequency f of streaked NEA detections per unit area

of sky per unit time (equivalently, per survey image). The posterior probability distribution of f (assuming a constant prior) is given by an appropriately-normalized Poisson distribution:

$$P(f) = \frac{(NC)^{n+1}}{\Gamma(n+1)} f^n \exp(-NCf) \tag{3.5}$$

where N is the total number of images searched for streaked NEAs, n is the number of detected streaked NEAs, C is the completeness (true positive rate) of the PTF streak detection system as a whole, and $\Gamma(\ldots)$ is the gamma function (which contributes to the normalization of the distribution).

Figures 5 and 11 indicate that the completeness C depends on which volume in magnitude, length, and orientation space under consideration, as well as the separate efficincies of sub-components like `findStreaks` and the machine classifier. For simplicity, in the following analysis we evaluate two separate values for C (0.5 and 0.7) but the most accurate values for C would in principle come from direct application of the completeness data in Figures 5, 11 and 12.

We apply Equation (3.5) to the image count N and streak count n within each of the thirteen bins in Figure 3.18. In particular, by numerical integration we compute the 16th and 84th percentiles of the resulting Poisson distributions, and plot these 1σ bounds as a function of distance from opposition in Figure 3.19. The estimates are 1–3 streaked NEA detections per 10^4 deg^2 of sky *near opposition*, dropping to about 1 or less beyond 40–50 deg from opposition. The images acquired by PTF from 2014-May-01 through 2014-Dec-01 represent 191,435 deg^2, and a total of 19 streaked NEAs (10 confirmed, 4 blindly recovered, 5 unconfirmed) were detected in these data. If the areal density of the streaks were independent of distance from opposition, this would correspond to a coarse estimate of $\sim$1 detected streak per 10^4 deg^2, in agreement with the radially-binned rates multiplied by the actualy radial distribution of images (which are mostly 40 deg or more from opposition).

3.6 Scaling laws for streaked asteroid detection

We here derive a quantitative 'figure of merit' (FoM) proportional to the average number of streaked asteroids detectable per unit time by a survey. The FoM will depend on a number of survey specifications including the:

- field of view Ω in deg^2,
- seeing width $\theta_{\rm PSF}$ in arcseconds,
- limiting magnitude of a point source $m_{\rm pnt}^{\rm lim}$, which is related to other parameters such as:
    - exposure time τ in seconds,
    - telescope aperture $A_{\rm tel}$ in meters,
    - sky background flux B (counts/second/arcsec^2),
- total duty cycle time $\tau_{\rm tot}$ per exposure, including integration, readout, and slew time.

Assume that the density of asteroids and their velocity distribution is independent of distance. The volumne of streaked asteroids detectable at any given time goes as $\Omega d_{\rm strk}^3$, where $d_{\rm strk}$ is the maximum distance at which an asteroid can be detected as a streak. The figure of merit (asteroids detectable per unit time) therefore scales as

$$\mathrm{FoM} \propto \frac{\Omega d_{\rm strk}^3}{\tau_{\rm tot}} \tag{3.6}$$

The maximum streaking distance $d_{\rm strk}$ is defined for some signal-to-noise ratio (SNR), typically SNR = 5, at which the object will be detected with an apparent magnitude $m_{\rm strk}^{\rm lim}$. When detected at this threshold SNR, the streaking object's distance $d_{\rm strk}$ must be closer than the distance $d_{\rm pnt}$ of a point source observed with the same SNR but fainter magnitude

$$m_{\rm pnt}^{\rm lim} = 5\log_{10}(d_{\rm pnt}/d_0) \tag{3.7}$$

where d_0 is a parameter encoding the asteroid's intrinsic flux (dependent on its size, albedo). Assume that $d_{\rm pnt}$ is much less than 1 AU, so that heliocentric distance does not factor into d_0. Also assume that the survey observes mostly around opposition, so that phase angle effects need not factor significantly into d_0.

Let F be the asteroid's flux and B the background flux. For a point source the SNR is

$$\mathrm{SNR}_{\rm pnt} = \frac{F/(4\pi d_{\rm pnt}^2)}{\sqrt{B}} \times \sqrt{\tau} \tag{3.8}$$

An asteroid's apparent angular rate $\dot{\theta}$ relates to its distance and perpendicular velocity component by $\dot{\theta} = v_\perp / d_{\rm strk}$. We thus define the streak-time

$$\tau_{\rm PSF} \equiv \theta_{\rm PSF}/\dot{\theta} = \theta_{\rm PSF} d_{\rm strk}/v_\perp \tag{3.9}$$

For streaking to occur, the exposure time τ must be significantly longer, i.e., $\tau \gg \tau_{\rm PSF}$. The signal-to-noise in each segment of length $\theta_{\rm PSF}$ along the streak's extent receives flux from the asteroid only for a duration $\tau_{\rm PSF}$, but the noise accumulates for the entire exposure τ. In terms of this quantity, the signal-to-noise in a streak *segment* of length $\theta_{\rm PSF}$ is

$$\mathrm{SNR}_{\rm seg} = \frac{F/(4\pi d_{\rm strk}^2)}{\sqrt{B}} \times \frac{\tau_{\rm PSF}}{\sqrt{\tau}} \quad (\text{for } \tau \gg \tau_{\rm PSF}) \tag{3.10}$$

The streak's *total SNR* is found by adding in quadrature the individual SNRs of the $\tau/\tau_{\rm PSF}$ such segments comprising it:

$$\begin{aligned} \mathrm{SNR}_{\rm strk} &= \left(\sum_{i=1}^{\tau/\tau_{\rm PSF}} \mathrm{SNR}_{{\rm seg},i}^2 \right)^{1/2} \\ &= \mathrm{SNR}_{\rm seg} \times \sqrt{\frac{\tau}{\tau_{\rm PSF}}} \\ &= \frac{F/(4\pi d_{\rm strk}^2)}{\sqrt{B}} \times \sqrt{\tau_{\rm PSF}} \quad (\text{for } \tau \gg \tau_{\rm PSF}) \end{aligned} \tag{3.11}$$

A useful approximation for the general case of any τ, with the appropriate limiting and intermediate behavior, is:

$$\mathrm{SNR}_{\mathrm{strk}} = \frac{F/(4\pi d_{\mathrm{strk}}^2)}{\sqrt{B}} \times \left(\frac{\tau}{1 + \tau/\tau_{\mathrm{PSF}}} \right)^{1/2} \tag{3.12}$$

Setting $\mathrm{SNR}_{\mathrm{pnt}} = \mathrm{SNR}_{\mathrm{strk}}$ using Equations (8) and (12),

$$\frac{d_{\mathrm{pnt}}^2}{d_{\mathrm{strk}}^2} = \sqrt{1 + \frac{\tau}{\tau_{\mathrm{PSF}}}} \tag{3.13}$$

Hence, the limiting magnitude for a streak, in terms of the limiting magnitude for a point source with the same SNR, is

$$\begin{aligned} m_{\mathrm{strk}}^{\mathrm{lim}} &= m_{\mathrm{pnt}}^{\mathrm{lim}} - 2.5 \log_{10}(d_{\mathrm{pnt}}^2/d_{\mathrm{strk}}^2) \\ &= m_{\mathrm{pnt}}^{\mathrm{lim}} - 1.25 \log_{10}(1 + \tau/\tau_{\mathrm{PSF}}) \end{aligned} \tag{3.14}$$

Equation (3.14) approximates the detection efficiency drop-off (limiting magnitude vs. length) seen in Figures 5 and 11.

If we once again consider the 'significantly-streaked' limit ($\tau \gg \tau_{\mathrm{PSF}}$) and use Equations (7), (9), and (13), we find

$$d_{\mathrm{strk}}^3 = \frac{d_0^4 \theta_{\mathrm{PSF}}}{v_\perp \tau} \times 10^{0.8 m_{\mathrm{pnt}}^{\mathrm{lim}}} \qquad (\text{for } \tau \gg \tau_{\mathrm{PSF}}) \tag{3.15}$$

Again assuming that the asteroids' size and albedo (represented by d_0) and velocities ($v_\perp$) are constant with geocentric distance, the figure of merit (Equation 3.6) for comparing surveys then takes the form

$$\mathrm{FoM} \propto \frac{\Omega \theta_{\mathrm{PSF}}}{\tau_{\mathrm{tot}} \tau} \times 10^{0.8 m_{\mathrm{pnt}}^{\mathrm{lim}}} \tag{3.16}$$

Approximate FoM values are listed in Table 3.5 relative to PTF. Any other survey's FoM can be computed by normalizing its limiting magnitude as $m_{\mathrm{pnt}}^{\mathrm{lim}} \rightarrow m_{\mathrm{pnt}}^{\mathrm{lim}} - 17.15$ and then applying Equation 3.16.

Table 3.5: Figure of merit comparison for surveys

survey	Ω (deg^2)	θ_{PSF} (arcsec)	τ (min)	τ_{tot} (min)	$m_{\mathrm{lim}}^{\mathrm{pnt}}$	FoM w.r.t. PTF
ZTF	47	2	0.5	0.75	20.4	37
PS1 or PS2	7	1.1	0.5	1	21.8	30
ATLAS	60	2.6	0.5	0.6	19.7	21
BlackGEM	22	1	1	1.2	20.7	4.7
Catalina	19	2.5	0.5	0.75	19.5	3.6
PTF	7.25	2.0	1	1.5	20.2	1

Chapter 4

Asteroid Lightcurves

4.1 Introduction

In this work we model an asteroid's apparent visual magnitude V (log flux) as

$$V = H + \delta + 5\log_{10}(r\Delta) - 2.5\log_{10}[\phi(\alpha)], \tag{4.1}$$

where H is the absolute magnitude (a constant), δ is a periodic variability term due to rotation (e.g., if the object is spinning and has some asymmetry in shape or albedo), r and Δ are the heliocentric and geocentric distances (in AU), and $\phi = \phi(\alpha)$ is the *phase function*, which varies with the solar phase angle α (the Sun-asteroid-Earth angle). When $\alpha = 0$ (i.e., at opposition), $\phi = 1$ by definition, while in general $0 < \phi < 1$ for $\alpha > 0$ (with ϕ decreasing as α increases).

A key feature of our approach is the simultaneous fitting of both the phase function ϕ and the rotation term δ. The detailed forms of ϕ and δ, as well as the algorithm underlying our fitting procedure, are motivated by a variety of prior work in this area, as described in the following sections.

4.1.1 Asteroid rotation

Building upon the work of [Kaasalainen et al. (2001)], [Hanuš & Ďurech (2012))] discuss the inversion of asteroid lightcurve data taken over several oppositions to obtain a 3D shape solution. The form of δ (cf. Equation [4.1]) in this case consists of

a large number of free parameters (several tens to hundreds). Results from inversion agree well with those from stellar occultations, adaptive optics imaging, and in-situ spacecraft imagery [Hanuš et al. (2013))]. Knowledge of the detailed irregular shapes of asteroids improves our ability to constrain models of their internal structure, as well the magnitude and timescale of spin and orbital evolution due to solar-radiation and thermal emission, including the Yarkovsky and YORP effects (see [Bottke et al. (2006)] and references therein).

A simpler model for δ—suitable for fitting to data sparser than that required for most inversion methods—is a Jacobi ellipsoid [Chandrasekhar (1969)] in its principal-axis spin state. The lightcurve of such an ellipsoid is a double-peaked sinusoid, given by a simple expression depending solely (assuming constant surface albedo) on the axes ratio, and angle between the line of sight and spin axis. The fitted amplitude thus yields a lower-bound elongation estimate for the asteroid.

The predicted distribution of the rotation frequencies of a collisionally-equilibrated system of particles has long been claimed to be a Maxwellian function [Salo (1987)], which—as reviewed by [Pravec et al. (2002)]—very well approximates the observed distribution of several hundred of the brightest (∼40-km or larger) asteroids, but breaks down for smaller objects, among which an excess of slow and fast rotators appear to exist. [Steinberg & Sari (2015)] more recently argue that collision instead leads to a Lévy distribution, and that a significant primordial spin component remains in the present observed population. Some studies that have examined the spin distribution of small objects are [Pravec et al. (2008)], [Polishook & Brosch (2009)], the Thousand Asteroid Lightcurve Survey [Masiero et al. (2009)], and two brief observing runs conducted within the PTF survey ([Polishook et al. (2012)]; [Chang et al. (2014a)]).

[Warner et al. (2009)] describe the Lightcurve Database (LCDB), which compiles several thousand densely-sampled lightcurves of asteroids targeted by dedicated observing teams. Lightcurves in the LCDB have the following features:

1. LCDB lightcurves' dense sampling generally permits fitting of Fourier series with many harmonic terms,

2. LCDB lightcurves are often sampled over the shortest time window necessary to measure the period, and therefore generally do not require large or uncertain corrections due to phase angle effects,

3. LCDB lightcurves' fitted periods are assigned integer quality codes by a human reviewer (from 1 = poor to 3 = confident).

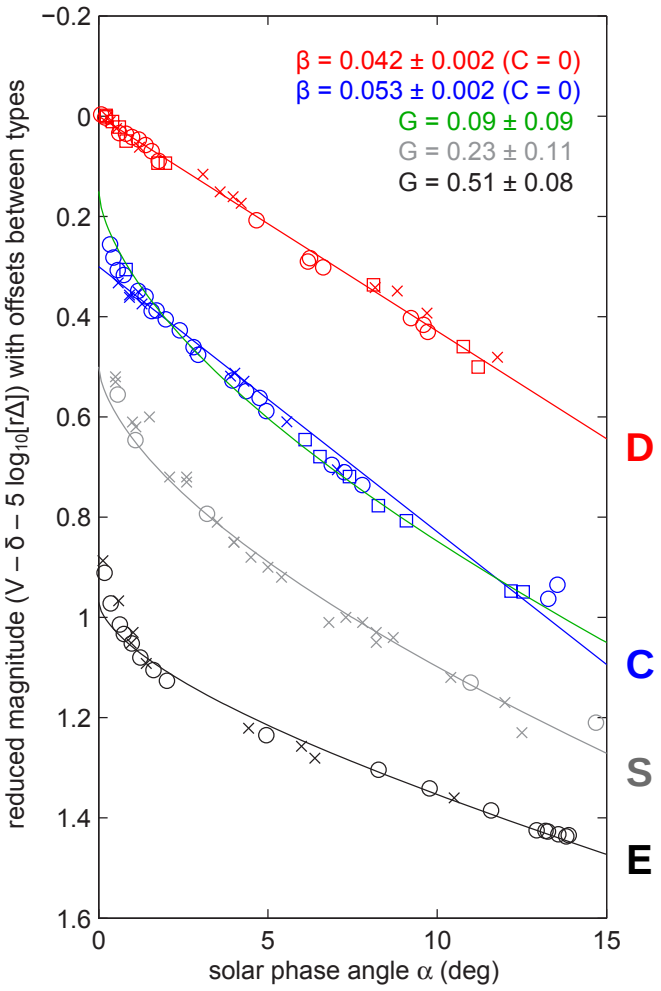
All three of the above features are either impractical or infeasible when the set of lightcurves is very large and the data sparsely sampled, as is the case for PTF. In this work we adopt the following modified approaches when fitting lightcurves:

1. We truncate the rotation curve's Fourier-series fit after the 2nd harmonic, a simplification broadly justified by [Harris et al. (2014)] and the assumption of an ellipsoidal shape (cf. Section 4.3.1.2),

2. We simultaneously fit a phase-function model with the rotational part,

3. We use a machine-learned classifier to objectively aid in estimating the validity of each fitted period. The classifier is trained using all fitted lightcurves that have previously (and confidently) measured LCDB periods and takes into account the accuracy with which the true period was retrieved along with 20 lightcurve metrics (fitted period, amplitude, ratio of peaks, χ^2 per degree of freedom of fit, number of data points, and more).

Use of a machine classifier in asteroid lightcurve period quality assessment is entirely novel and inspired in part by work done by PTF collaborators in extragalactic transient science [Bloom et al. (2012)] and variable star science ([Masci et al. (2014)]; [Miller et al. (2014)]), as well as Waszczak et al. (in prep)'s work on detection techniques for streaking NEOs. Among the advantages of using a machine-classified quality score is that, via cross-validation with the known-period sample, one estimates the completeness and contamination, i.e., the true-positive and false-positive rates with respect to identifying an accurately-fit period, as a function of, e.g., the period, amplitude, etc. The resulting true- and false-positive rates may then be used to de-bias the classifier-filtered period distribution.



Symbol		Asteroid	Reference
D	○	588 Achilles	Shevchenko et al. 2012
	×	884 Priamus	Shevchenko et al. 2012
	□	1143 Odysseus	Shevchenko et al. 2012
C	○	24 Themis	Harris et al. 1989a
	×	165 Loreley	Harris et al. 1992
	□	211 Isolda	Harris & Young 1989
S	○	20 Massalia	Gehrels 1956
	×	249 Amphitrite	Lupishko et al. 1981
E	○	44 Nysa	Harris et al. 1989b
	×	64 Angelina	Harris et al. 1989b

Figure 4.1: Phase curves (from the literature) containing densely-sampled, rotation-corrected photometry of asteroids in four taxonomic classes. Colored lines are our original fits to the data using various single-parameter ϕ models (cf. Section 4.3.2).

4.1.2 Asteroid phase functions

The analytic phase function of an ideal Lambertian-scattering sphere fits well to featureless, atmospheric planets like Venus, but quite poorly to airless bodies (see Figure 3.9 of [Seager (2010)] for a comparison). In later sections we describe several ϕ models that have been derived for (or empirically fit to) asteroids. Qualitatively, asteroids show an approximately linearly decreasing ϕ out to $\alpha \approx 100$ deg, modified by a surge (increase in slope) at low phase angles ($\alpha \lesssim 5$ deg), known as the *opposition effect* (see Figure 4.1).

Early work (e.g. [Bowell et al. (1989)] and refs. therein) on a small sample of well-observed asteroids, suggested that different asteroid spectral types display distinct behavior in ϕ. Figure 4.1 compares example phase curve data for D, C, S and E types[1], incorporating photometry from various sources. We emphasize the fact that all of the data points in Figure 4.1 have been corrected for rotational modulation (the δ in Equation [4.1]) through dense sampling of each asteroid's lightcurve at each phase angle (equivalently, each epoch).

Using a large corpus of low-precision photometry from the MPC[2], Oszkiewicz et al. ([Oszkiewicz et al. (2011)], [Oszkiewicz et al. (2012)]) showed that a fitted parameter of one particular ϕ model correlates well with an asteroid's SDSS visible color. While they were unable to correct for rotational variation (δ-term in Equation [4.1]), the Oszkiewicz et al. work nevertheless demonstrates a solid trend between ϕ and a compositional attribute (color).

These prior works motivate several defining aspects of this work's phase-function analysis:

1. We fit multiple phase function models to each lightcurve, both for compatibility with the literature and to explore how the fitted parameters are related,

2. We simultaneously fit the rotational component with the phase-function part,

[1] [Bus et al. (2002)] review these and other asteroid taxonomic classes, which are defined on the basis of low-resolution ($R \approx 100$) visible reflectance spectra.

[2] IAU Minor Planet Center, http://minorplanetcenter.nethttp://minorplanetcenter.net

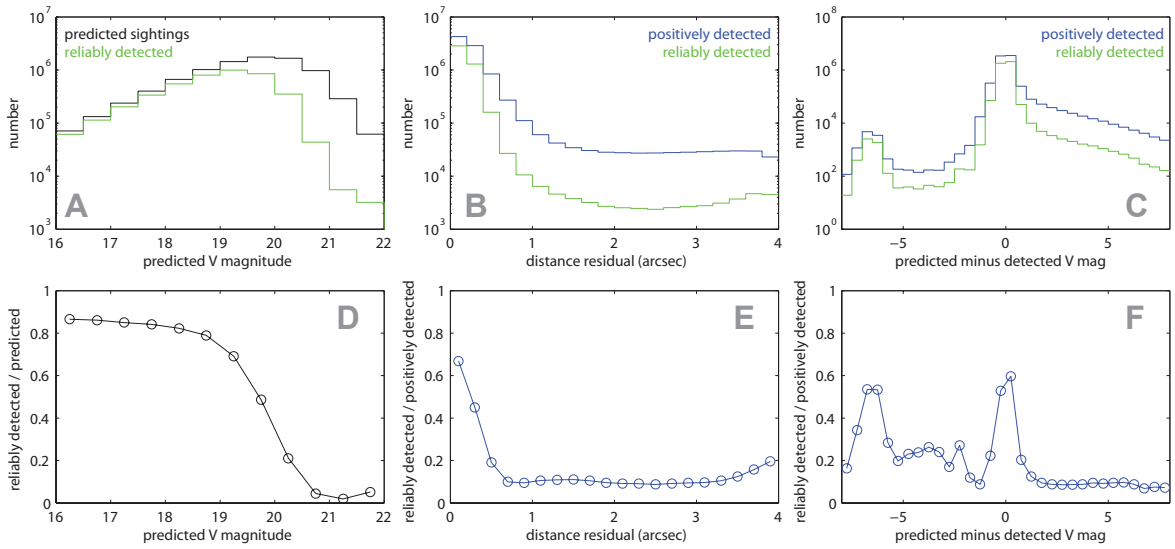


Figure 4.2: Comparison of predicted asteroid sightings against positive and 'reliable' asteroid detections. We define a 'reliable' detection as any positive detection which (1) lacks any *catalogued* background sources within a 4″ radius, (2) has a calibrated magnitude uncertainty of less than 0.1 mag, (3) lacks any processing warning flags. As suggested by the middle and right column of plots, this definition of 'reliable' still contains some small contamination (at the <1% level) from uncatalogued background sources and/or noise, as indicated by detections with distance residuals greater than ∼1 arcsecond or magnitude residuals of greater than ∼1 mag. In panel D, the less than 100% completeness at the bright end reflects the non-negligible probability that any asteroid will fall within 4″ of a catalogued background source (regardless of the magnitude of either the asteroid or the background source).

3. We introduce a single colorimetric index for quantifying C-type vs. S-type taxonomic classification, based on the compilation of several visible-band-color asteroid datasets (see Appendix), and examine the variation in phase-function parameters as a function of this color index.

4.2 Observations

4.2.1 Overview of the PTF survey

The *Palomar Transient Factory*[3] (PTF) is a synoptic survey designed primarily to discover extragalactic transients ([Law et al. (2009)]; [Rau et al. (2009)]). The PTF camera, mounted on Palomar Observatory's 1.2-m Oschin Schmidt Telescope, uses 11 CCDs (each 2K × 4K) to image 7.3 deg^2 of sky at a time at 1.0″/pixel resolution. Most exposures (∼85%) use a Mould-R filter[4] (hereafter "R"). The remaining broadband images acquired use a Gunn g-band filter. Nearly all broadband PTF images are 60-second integrations, regardless of filter. About 15% of nights (near full moon) are devoted to a narrowband (Hα) imaging survey of the full Northern Sky.

Science operations began in March 2009, with a nominal one- to five-day cadence for supernova discovery and typical twice-per-night imaging of fields. Median seeing is 2″ with a limiting magnitude $R \approx 20.5$ (for 5σ point-source detections), while dark conditions routinely yield $R \approx 21.0$ [Law et al. (2010)].

The PTF survey is ongoing and expected to continue through mid-2016. In January 2013 the PTF project formally entered a second phase called the *intermediate PTF* ('iPTF'; [Kulkarni (2013)]). In this chapter we simply use 'PTF' to mean the entire survey, from 2009 through the present (2015). The iPTF program accommodates more varied 'sub-surveys' as opposed to a predominantly extragalactic program, including variable star and solar system science. Images are still acquired with the same telescope/camera/filters with 60s exposures, and are processed by the same

[3] http://ptf.caltech.eduhttp://ptf.caltech.edu

[4] The Mould-R filter is very similar to the SDSS-r filter; see [Ofek et al. (2012a)] for its transmission curve.

reduction pipeline.

[Laher et al. (2014)] describe the PTF data reduction and archiving pipelines, hosted at the Infrared Processing and Analysis Center (IPAC) at Caltech. Processing at IPAC includes bias and flat-field corrections, astrometric calibration against UCAC3 [Zacharias (2010)], astrometric verification against 2MASS [Skrutskie et al. (2006)], creation of source catalogs with Source Extractor [Bertin and Arnouts (1996)], and production of reference images (stacks of ∼20–30 PTF images that reach $V \approx 22$).

Ofek et al. ([Ofek et al. (2012a)], [Ofek et al. (2012b)]) describe the PTF survey's absolute photometric calibration method, which relies on source matching with SDSS DR7 [Abazajian et al. (2009)], and thus requires at least partial overlap of PTF with SDSS each night. A separate, *relative* photometric calibration (based on lightcurves of non-variable field stars) also exists for PTF data and is described by [Levitan et al. (2011)] and in the appendix of [Ofek et al. (2011)]. In this work we utilize all R-band and g-band PTF data accumulated from the survey's start (March 2009) through July 2014. The asteroid magnitudes reported in this work use relative photometric zeropoints when available (which as of this writing applies to ∼85% of PTF images) and absolute photometric zeropoints otherwise.

The PTF's robotic survey program and processing pipeline, as well as our data aggregation and analysis in this work, make use of many functions from the MATLAB package for astronomy and astrophysics [Ofek et al. (2014)].

4.2.2 This work's data set

[Waszczak et al. (2013b)] used a custom spatial indexing algorithm to search the set of all PTF single-epoch transient detections (through July 2012) for detections of all asteroids with orbits known as of August 2012. That search procedure first generated uniformly-spaced ephemerides for each asteroid using JPL's online service (HORIZONS; [Giorgini et al. (1996)]). Each asteroid's ephemeris defines a 3D-curve (two sky coordinates plus one time); the intersection of each curve with the 3D kd-

Table 4.1: Description of the PTF asteroid database. Includes PTF data acquired from March 2009 through July 2014, excluding Hα survey data.

table	# rows	example columns (not necessarily comprehensive)
PTF tiles	11,169	R.A., Dec., tile ID
exposures	304,982	epoch, filter, exposure time, absolute photometric zeropoint, tile ID, exposure ID
CCD images	3,305,426	CCD ID, corners RA & Dec, seeing, limiting mag., relative phot. zeropoint, # of sources, exposure ID, image ID
asteroids	401,810	name, orbital elements, color data (e.g., SDSS), IR data (e.g., WISE), known rotation period, asteroid ID (number)
predicted sightings	17,929,274	R.A., Dec., rates, helio- & geocentric range, phase & elong. angle, pred. V mag., image ID, asteroid ID, prediction ID
positive detections	8,842,305	R.A., Dec., instrumental mag., local zeropoint, shape data, quality flags, prediction ID, lightcurve ID, detection ID
reliable detections[1]	4,392,395	detection ID
lightcurves[2]	587,466	# of constituent detections, filter, opposition year, median mag., asteroid ID, lightcurve ID
lightcurve fits[3]	54,296	fitted lightcurve parameters, human-assigned quality code, machine-classified quality index, lightcurve ID, fit ID
reliable-period fits[4]	9,033	fit ID
reliable-G_{12} fits[5]	3,902	fit ID

[1]'Reliable' detections are those free from possible background-source or bright star contamination, magnitude errors > 0.1 mag, and certain SExtractor flags.

[2]A lightcurve is here defined as a set of positive detections of a given asteroid in a single filter and opposition.

[3]Lightcurve fits only exist for lightcurves which contain at least 20 reliable detections and converged to a solution during the lightcurve-fitting process.

[4]Fits have reliable rotation periods if a human screener labels the period reliable *and* the machine classifier rates it above a certain quality threshold (see text).

[5]Fits have reliable G_{12} phase-function parameter if (1) amplitude < 0.1 mag *or* period is reliable, (2) fit has sufficient phase angle coverage (see section 4.6.3).

tree of transient detections was then computed and positive detections within a 4″ matching radius saved.

In this work we use a modified version of the [Waszczak et al. (2013b)] algorithm. The updates/changes are as follows.

Firstly, in terms of content, we now search all PTF (R and g-band) data from 01-March-2009 through 18-July-2014 for all *numbered* asteroids as of 12-July-2014 (401,810 objects). We now exclude unnumbered objects as the positional uncertainty of these objects can be very large, and as they tend to be very faint their lightcurves will not in general be of high quality.

Secondly, in place of a single-step matching of a 3D transient-detection kd-tree against 3D ephemeris curves, we now divide the search into two main steps. We first perform a 2D spatial matching that exploits the natural indexing of PTF exposures into tiles (i.e., the grid of evenly spaced boresights or 'fields' on the sky). Each 2D ephemeris curve's intersection with the 2D PTF survey footprint is computed, the object's position cubically-interpolated to all epochs of exposures possibly containing the object, and the object's precisely-computed positon is then compared to the precise image boundaries of candidate exposures. Matching of predicted positions against actual detections takes place subsequently as source catalogs are then loaded into memory (as needed and in parallel). This method is faster than the original [Waszczak et al. (2013b)] method and enables separate logging of predicted and positive detections.

The results of the known-asteroid search, as well as the derived lightcurve data (described later) are stored in a relational database, the size and contents of which are summarized in Table 4.1. Out of ∼18 million predicted single-epoch asteroid sightings (including predicted magnitudes as dim as $V \approx 23$, well below PTF's sensitivity), there were 8.8 million positive detections (within a 4″ radius). Of these, we define 4.3 million detections as 'reliable' as they (1) lack any *catalogued* background sources within the 4″ radius, (2) have a calibrated magnitude uncertainty of less than 0.1 mag, (3) lack any processing flags indicative of contamination. Figure 4.2 compares predicted, positive and 'reliable' detections; the middle and right panels of Figure 4.2

show that our definition of 'reliable' seems to include a small fraction of likely bad observations (<1% contamination, note the vertical log scale), namely those which have distance residuals greater than $\sim 1''$ or magnitude residuals greater than ~ 1 mag. Because these reliable detections are the subset of observations which we input into our lightcurve fitting model (Section 4.3.4), the fitting algorithm includes logic designed to remove isolated data points that have very large residuals, either with respect to the median lightcurve value or relative to their uncertainty.

4.3 Lightcurve model

Equation (4.1) presents the overall form and notation of our asteroid lightcurve model. In this section we describe the detailed parameterization and assumptions of the model.

4.3.1 Rotation component

4.3.1.1 Intra-opposition constraint

The most important parameter in the rotation component (the δ in Equation [4.1]) is the synodic spin period P, a constant which satisfies

$$\delta(\tau) = \delta(\tau + nP), \tag{4.2}$$

where $\tau \equiv t - \Delta/c$ is the light-time-corrected observation timestamp, $\Delta = \Delta(t)$ is the asteroid's geocentric distance, c is the speed of light, and n is any integer satisfying

$$|n| \ll P_{\rm orb}/P, \tag{4.3}$$

where $P_{\rm orb}$ is the synodic *orbital* period,

$$P_{\rm orb} = \left(\frac{1}{\rm yr} - \frac{1}{T_{\rm orb}}\right)^{-1} = \left(\frac{1}{\rm yr} - \frac{\sqrt{GM_\odot}}{2\pi a_{\rm orb}^{3/2}}\right)^{-1}, \tag{4.4}$$

where $T_{\rm orb}$ is the asteroid's *sidereal* orbital period and $a_{\rm orb}$ is its orbital semi-major axis (related by Kepler's third law). $P_{\rm orb}$ is the time elapsed between the asteroid's consecutive oppositions. Pursuant to this restriction, we constrain each δ solution using observations from within the same opposition—i.e., for most asteroids, within a 1.1- to 1.6-year interval centered on the date of locally minimally observed α.

The intra-opposition restriction is important given that our data set (described in the next section) spans ∼5 years. For an asteroid with a zero inclination circular orbit and spin axis perpendicular to its orbital plane, we can relax Equation (4.3) to allow n to be any integer, in which case δ can be constrained using observations spanning many years. In general however, Equation (4.2) must be modified to accommodate a varying viewing geometry with respect to the spin axis:

$$\delta(\tau) = F(\tau)\delta(\tau + nP), \tag{4.5}$$

where F is some unknown periodic function satisfying $F(t) = F(t + mT_{\rm orb})$, where m is any integer and $T_{\rm orb}$ is the sidereal orbital period. Provided the amplitude of F is not large relative to that of δ, and provided the spin vector is not changing with respect to the orbital plane (i.e., precessing[5]) on a timescale comparable to $P_{\rm orb}$, we are justified in assuming Equation (4.2) (with the Equation [4.3] restriction) applies.

4.3.1.2 Second-order Fourier series

Any δ satisfying Equation (4.2) can be approximated to arbitrary precision using a Fourier series. [Harris et al. (2014)] discuss why, from a geometric standpoint, the second harmonic tends to dominate an asteroid's fitted δ. As noted earlier (section 4.1.1), most large asteroids approximately resemble triaxial prolate ellipsoids (e.g., Jacobi ellipsoids), having equatorial axis ratios of at most ∼3:1 (corresponding to a $\delta_{\rm max} - \delta_{\rm min}$ amplitude of ∼1.2 mag). For less extreme axis ratios (specifically, those producing a ∼0.4 mag or smaller second-harmonic amplitude), other harmonics related to shape or albedo asymmetries may contribute comparable coefficients to the

[5]Principal-axis rotation (a stable equilibrium state) is assumed for most planetary bodies. [Burns & Safronov (1973)] discuss the relevant timescales of spin evolution.

Fourier approximation of δ.

The PTF survey program has—on a few rare occasions—conducted high-cadence ($\sim$10-minute spaced) observations of low ecliptic latitude fields. These runs produced a set of $\sim$1,000 densely-sampled main-belt asteroid rotation curves, which have already been analyzed and published ([Polishook et al. (2012)]; [Chang et al. (2014a)]). These high-cadence "pilot studies" are relevant to our present work in that they demonstrate (1) the quality of the PTF survey's photometric calibration for asteroids with unambiguously valid δ solutions, and (2) the above-described prevalence of a dominant second-harmonic in most of the objects sampled.

Following these pilot studies, we adopt a second-order Fourier series model:

$$\delta \equiv \sum_{k=1,2} A_{1,k} \sin\left(\frac{2\pi k\tau}{P}\right) + A_{2,k} \cos\left(\frac{2\pi k\tau}{P}\right), \tag{4.6}$$

where τ is the light-time corrected epoch (cf. Equation [4.2]). In the pilot studies, most of the fitted δ solutions qualitatively resemble a simple sine or cosine function. Such a solution can be represented by *either* a:

1. first harmonic with period $P = P_1$ (with $A_{i,1} \neq 0$ and $A_{i,2} = 0$), *or*

2. second harmonic of period $P = \frac{1}{2}P_1$ (with $A_{i,1} = 0$ and $A_{i,2} \neq 0$).

Given the prolate ellipsoid model, choice (2) is more realistic and hence preferred. However, again recognizing that other harmonics can have a non-negligible contribution, in fitting δ to our lightcurve sample we allow the first-harmonic coefficients $A_{i,1}$ to be non-zero, but introduce logic into the fitting algorithm (cf. Section 4.4) which checks for double-period solutions satisfying certain criteria and iterates accordingly.

4.3.2 Phase-function component

In this work we simultaneously fit each lightcurve's phase function ϕ along with its rotation curve δ (cf. Equation [4.1]). This approach is intermediate in complexity between some of the simpler, two-parameter (δ-neglecting) models that have been

applied to very large data sets (e.g., [Williams (2012)]; [Oszkiewicz et al. (2012)]), and the more complex, shape plus pole-orientation models ([Kaasalainen (2004)]; [Cellino et al. (2009)]; [Hanuš & Ďurech (2012))]) which can involve tens of parameters and require data spanning multiple oppositions.

Regarding the former class of models, we note that there is a formal statistical problem associated with neglecting δ when fitting ϕ. If modeling the observations M by $V' \equiv V - \delta = H + 5\log_{10}(r\Delta) - 2.5\log_{10}(\phi)$, then the distribution of residuals $M - V'$ is *not* Gaussian. Assuming δ is a sinusoid with amplitude A, for observations M sampling the lightcurve at random times, the residual probability density function $p = p(M - V')$ has a local minimum value $p_{\min}$ at $M - V' = 0$ and maximum value $p_{\max}$ near $M - V' = \pm A$. Thus p is bimodal and roughly bowl-shaped—not at all Gaussian-shaped. The uncertainty in ϕ produced by a standard χ^2 minimization—which assumes Gaussian-distributed errors—is thus inaccurate. However, since p is symmetric about $M - V' = 0$, for densely-sampled data the fitted phase function ϕ remains unaffected by neglecting δ; in such a case the only effect is an underestimated uncertainty.

We obtain three separate fits for each lightcurve, each using a different phase-function (ϕ) and allowing for unique solutions for H and δ in Equation (4.1). The three phase-function models are:

1. the two-parameter model of [Shevchenko (1997)],
2. the one-parameter G model [Bowell et al. (1989)],
3. the one-parameter G_{12} model [Muinonen et al. (2010))].

In this section we review and motivate the application of each of these ϕ models.

4.3.2.1 Two-parameter Shevchenko model

[Shevchenko (1997)] introduced a phase function dependent on two parameters; in terms of Equation (4.1) the model is[6]

[6]In Shevchenko's original notation, β is denoted b and C is denoted a. Moreover, in the original notation, $\phi(0) = -a$; we here added a constant term $+a$ to make $\phi(0) = 1$, following convention

$$-2.5\log_{10}[\phi(\alpha)] \equiv \beta\alpha - C\frac{\alpha}{1+\alpha}, \tag{4.7}$$

where β has units of mag/deg and C is the amplitude of the opposition surge (units of mag). This model was subsequently considered in-depth by [Belskaya & Shevchenko (2000)], hereafter B&S, who compiled the most complete (to date) set of high-precision, targeted phase curve observations of main-belt asteroids from various data sets spanning several decades.

Though in practice Shevchenko's model is the least commonly used phase function out of the three we consider, it is by far the simplest to express mathematically, and is the only model for ϕ whose parameters have linear dependence in Equation (4.1).

Furthermore, this model's parameters are the most straightforward to associate with physical asteroid properties. B&S highlighted a robust relationship between an asteroid's (β, C) phase-function parameters and its geometric albedo[7]. As we later explore a similar relationship in the present work, we here review the basis of this observation.

The geometric albedo p_V is formally *defined* in terms of the phase function ϕ:

$$p_V \equiv \frac{A_{\rm bond}}{2}\left(\int_0^\pi \phi(\alpha)\sin(\alpha)\,d\alpha\right)^{-1} \equiv \frac{A_{\rm bond}}{q}, \tag{4.8}$$

where $A_{\rm bond}$ is the (visible) bond albedo, defined as the total visible light energy reflected or scattered by the asteroid (in all directions) divided by the total visible light energy incident upon the asteroid (from the Sun). We also here define the phase integral q.

B&S showed that, in the range of β observed from S-type to C-type asteroids, β and C are empirically correlated, in a relation that we approximate here as

$$C \approx (0.9\text{ mag}) - (17\text{ deg})\beta \;\text{ for }\; 0.03 < \frac{\beta}{\text{mag/deg}} < 0.05. \tag{4.9}$$

Using Equation (4.9) to substitute for C in Equation (4.7), inserting the result into

with other phase functions.

[7]Also known as the *visible* albedo or the *physical* albedo.

Equation (4.8) and numerically evaluating the integral gives

$$p_V \approx A_{\text{bond}} \left(0.4 - \frac{2.2\beta}{\text{mag/deg}}\right) \quad \text{for} \quad 0.03 < \frac{\beta}{\text{mag/deg}} < 0.05. \tag{4.10}$$

B&S saw a negative correlation between p_V and β in the data[8], consistent with Equation (4.10) *only if* either A_{bond} is assumed constant among different asteroid types (not a reasonable assumption) *or* if A_{bond} negatively correlates with β, which B&S did not explicitly show.

The bond albedo A_{bond} can be thought of as an intrinsic, bulk-compositional characteristic of an asteroid's surface[9], much like an asteroid's color, whereas β and C relate (in part) to the textural, particulate, and macroscopic roughness of the asteroid's surface. B&S and other authors separately associate β with the *shadow-hiding* effect and C with the *coherent backscatter* effect. Both of these physical phenomena are understood from a theoretical standpoint (e.g., [Helfenstein & Veverka (1989)]; [Hapke (2012))]) to be functions of A_{bond}, with β negatively related to A_{bond} and C positively related. This is consistent with Equation (4.9), and renders Equation (4.10) consistent with B&S's noted p_V-vs.-β correlation. Other properties such as particle size, particle geometry and regolith porosity also have predicted (and laboratory-measured) contributions to the observed phase function ([Hapke (2012))] and refs. therein); these properties can conceivably vary independently of A_{bond}.

In short, our interpretation of the S-type and C-type asteroid data reviewed by B&S is that a compositional indicator (A_{bond}) correlates with indicators of two independent phenomena (β and C) that contribute to how light scatters from an asteroid's surface. This statement intentionally makes no mention of p_V, since Equation (4.8) tells us p_V by definition varies with β (in a non-obvious way) and with A_{bond}, the latter being a more basic compositional attribute.

As stated above, the phase function can be related to properties other than A_{bond},

[8] B&S actually stated the correlation in terms of $\log p_V$ vs. β, though the range in β is sufficiently small that p_V vs. β is essentially valid as well.

[9] More accurately, the single-scattering albedo w, which is the analog of A_{bond} for a "point-source" particle, more fundamentally embodies this bulk-compositional attribute. [Hapke (2012))] details how A_{bond} is solely a function of w for an asteroid whose surface consists of isotropic scatterers; we here use A_{bond} as a proxy for w.

such as regolith porosity. Many of these other properties in theory and experiment contribute to effects involving *multiply*-scattered light, and therefore do not alter the effect of shadow-hiding (β-term in Equation [4.7]), which is dominated by *singly*-scattered light [Hapke (2012))]. In contrast, the coherent backscatter effect (C-term) *does* involve multiply-scattered light. B&S saw non-monotonic behavior in C as a function of p_V when including the rarer, high-p_V E-type asteroids in the same plot as C and S types. E types do conform however to the same negative monotonic trend in p_V-vs.-β satisfied by the C and S types, consistent with the hypothesis that β is adequately expressed as a function of A_{bond} alone, yet E types have a lower-than-predicted C value based on extrapolation of Equation (4.9).

One possibility is that Equation (4.9) is not valid for all asteroids, but must be replaced by some unknown non-monotonic relationship, possibly because C depends non-monotonically on A_{bond} and/or has comparable dependence on other properties (e.g., porosity or grain size). Assuming Equation (4.7) is a sufficiently general model for ϕ, and lacking knowledge of a good model for C, it follows that β and C should in practice always be fit separately. Another possibility is that Equation (4.7) is an incorrect or incomplete model, however B&S described no instances wherein their model was unable to adequately fit the data for a particular asteroid or class of asteroids.

4.3.2.2 Lumme-Bowell G model

The next phase function model we consider is the Lumme-Bowell model [Bowell et al. (1989)], also known as the (H,G) or IAU phase function:

$$\begin{cases} \phi \equiv (1-G)\phi_1 + G\phi_2 \\ \phi_1 \equiv \exp(-3.33\tan^{0.63}[\alpha/2]) \\ \phi_2 \equiv \exp(-1.87\tan^{1.22}[\alpha/2]) \end{cases} \tag{4.11}$$

Like Shevchenko's model, this model includes two terms (the basis functions ϕ_1 and ϕ_2) representing two physically-distinct contributions to the observed ϕ. As

detailed in [Bowell et al. (1989)], this model is *semi*-empirical in that it was derived from basic principles of radiative transfer theory with certain assumptions, and at various stages tailored to match existing laboratory and astronomical observations. That the two basis functions' coefficients are related to a single parameter G bears resemblance to the β-vs.-C correlation described by Equation (4.9).

[Marsden (1986)] marked the IAU's adoption of this phase function as a standard model for predicting an asteroid's brightness. Since then this model has seen widespread application, and is often used with the assumption $G = 0.15$ (e.g., in the ephemeris computation services offered by the MPC and JPL). [Harris & Young (1988)] present mean values of G for several of the major asteroid taxonomic classes (based on a sample of ∼80 asteroids), with $G = 0.15$ being an average between the C types ($G \approx 0.08$) and the S types ($G \approx 0.23$). The G-model fails to accurately fit the rarer D types (which have linear phase curves) and E types (which have very sharp opposition spikes), whereas the Shevchenko model can properly accommodate these rarer types.

Use of the Lumme-Bowell ϕ in our lightcurve model (Equation [4.1]) introduces a second non-linear parameter (G) into the model, the period P being the other non-linear parameter. This complicates the fitting algorithm somewhat, as described in Section 4.4.

4.3.2.3 Muinonen et al. G_{12} model

The third phase function model we consider, introduced by [Muinonen et al. (2010))], bears resemblance to the G-model but includes a second free parameter and a third basis function:

$$\phi \equiv G_1\phi_1 + G_2\phi_2 + (1 - G_1 - G_2)\phi_3 \tag{4.12}$$

As opposed to the analytic trigonometric basis functions of the G-model, here ϕ_1, ϕ_2 and ϕ_3 (all functions of α alone) are defined in terms of cubic splines (see [Muinonen et al. (2010))] for the exact numerical definitions). Assuming the coeffi-

cients G_1 and G_2 are constrained independently, these basis functions were designed to provide the most accurate fits to the phase functions of all major asteroid taxonomic types, including the rarer D types and E types.

For situations where fitting G_1 and G_2 separately is infeasible, [Muinonen et al. (2010))] specialized their above model to make it a function of a single parameter, G_{12}, which parameterizes G_1 and G_2 using piecewise functions:

$$G_1 = \begin{cases} 0.7527G_{12} + 0.06164 & \text{if } G_{12} < 0.2; \\ 0.9529G_{12} + 0.02162 & \text{otherwise;} \end{cases} \tag{4.13}$$

$$G_2 = \begin{cases} -0.9612G_{12} + 0.6270 & \text{if } G_{12} < 0.2; \\ -0.6125G_{12} + 0.5572 & \text{otherwise;} \end{cases}$$

In this work we use this single-parameter G_{12} form of the Muinonen et al. model, making it analogous to the G-model in terms of implementation, including the complication associated with a non-linear parameter.

4.3.2.4 Multi-parameter Hapke model

Just as we commented on the more rigorous means of fitting a rotation curve via 3D shape modeling with multi-opposition data, for completeness we note that a more rigorous model (than the three presented above) exists for phase functions. Given better-sampled lightcurves and more computational power, future modeling of large photometric datasets would benefit from applying the more theoretically-motivated model of [Hapke (2012))], an abbreviated form of which is

$$\phi = \frac{B_C K}{p_V} \left[\left(\frac{w}{8}(B_S g - 1) + \frac{r_0 - r_0^2}{2} \right) h + \frac{2}{3} r_0^2 \phi_L \right] \tag{4.14}$$

Here w is the single-scattering albedo (cf. Footnote 9), of which r_0 is solely a function. The remaining factors all are functions of phase angle (α). Each opposition-surge term (B_S and B_C) has two free parameters (width and amplitude). K depends on the mean topographic roughness (a function of one free parameter); g is the single-

scattering angular distribution function (typically includes one parameter); h is a function of α only; and ϕ_L is the phase function of an ideal Lambertian-scattering sphere (a simple function of α).

With its $\phi \propto p_V^{-1}$ dependence, the Hapke model (Equation [4.14]) can conveniently eliminate both p_V and H from the modeling process. Inserting Equation (4.14) into Equation (4.1), and using the common relation[10]

$$H = -5 \log_{10} \left(\frac{D\sqrt{p_V}}{1329 \text{ km}} \right), \tag{4.15}$$

where H is the absolute visual magnitude, D is the asteroid's effective diameter and 1329 km is a constant (set by the arbitrarily-defined magnitude of the Sun), produces a model with many physically meaningful parameters and free of both H and p_V.

4.4 Lightcurve-fitting algorithm

We solve Equation (4.1) using a custom *linear least squares* (LLSq) method. A basic review of LLSq can be found in [Hogg et al. (2010))]. Each fitted asteroid lightcurve contains $N_{\text{obs}} \geq 20$ observations, with measured apparent magnitudes m_i and measurement uncertainties σ_i. All instrumental magnitudes are elliptical aperture [Kron (1980)] measurements (SExtractor's `MAG_AUTO`) calibrated with a local zeropoint (i.e., the 'ZPVM' correction of [Ofek et al. (2012a)]). The uncertainties contain a Poisson-noise component (SExtractor's `MAGERR_AUTO`) as well as systematic error from the calibration. For images lacking a relative photometric solution, the relevant systematic error is the `APBSRMS` parameter in the PTF database; for images having a relative photometric solution, the systematic error is a combination of the `sysErr` and `zeroPointErr` database quantities (added in quadrature).

[10] Rather that attributing it to any specific author(s), we note that Equation (4.15) may be derived directly using Equation (4.8) and the following definition of the bond albedo, which we stated in words immediately after Equation (4.8):

$$A_{\text{bond}} \equiv \frac{\int_0^\pi 10^{-V(\alpha)/2.5} \sin(\alpha) d\alpha}{(10^{-M_{\text{Sun}}/2.5}/4\pi \text{AU}^2) \times \pi (D/2)^2}$$

where $V(\alpha) = H - 2.5 \log_{10} \phi(\alpha)$ is Equation (4.1) evaluated at $\delta = 0$ and $r = \Delta = 1$ AU.

In all cases, our model (Equation [4.1]) is non-linear in at least one parameter (the period P, or equivalently the frequency $f = 1/P$). We test $N_{\rm frq}$ evenly-spaced frequencies between $f = 0$ (infinite rotation period) and $f = 12$ day^{-1}, i.e., up to the $\sim$2-hour spin barrier.

Asteroids rotating *faster* than the $\sim$2-hour spin barrier are likely monolithic objects and—particularly if larger than $\sim$150 m—are interesting in their own right (cf. the discussion in [Pravec et al. (2002)]). However, given the apparent observed rarity of such super-fast rotators (SFRs) and the large interval in frequency space that must be searched to discover them; we impose 2 hours = 12 cycles per day as our upper limit on fitted frequency in order to make computational time reasonable without sacrificing sensitivity to the majority of asteroids' spin rates. [Chang et al. (2014a)] presents preliminary results of an independent, ongoing effort to use PTF data (or at least specific subsets thereof) to search for SFRs, with at least one SFR having been discovered and confirmed [Chang et al. (2014b)].

We use a frequency spacing $\Delta f = 1/(4\Delta t)$, where Δt is the time interval between the first and last observation in the lightcurve. Formally Δt can be as long as 1.1 to 1.6 yr for most asteroids (cf. Section 4.3.1.1); however the median value of Δt (among lightcurves that ultimately acquired fits) is $\sim$45 days, with 16$^{\rm th}$ and 84$^{\rm th}$ percentiles of 13 and 106 days, respectively.

In addition to the non-linear parameter f, the lightcurve model in general has $N_{\rm lin}$ linear parameters. We seek to solve the following tensor equation for X:

$$m_i = \sum_{j,k} L_{ijk} X_{jk} \quad \begin{cases} i = 1, 2, ..., N_{\rm obs} \\ j = 1, 2, ..., N_{\rm frq} \\ k = 1, 2, ..., N_{\rm lin} \end{cases} \tag{4.16}$$

where m_i is the $i^{\rm th}$ observation, L is the 'design matrix' (a 3D array of size $N_{\rm obs} \times N_{\rm frq} \times N_{\rm lin}$) and X is the linear-parameter matrix ($N_{\rm frq} \times N_{\rm lin}$) containing the linear-parameter solutions as a function of frequency.

4.4.1 Linear phase-function parameters

For the particular case wherein we use Shevchenko's model (Equation [4.7]) for the phase function ϕ, the design matrix is

$$L_{ij} = \begin{pmatrix} 1 \\ \sin(2\pi f_j \tau_i) \\ \cos(2\pi f_j \tau_i) \\ \sin(4\pi f_j \tau_i) \\ \cos(4\pi f_j \tau_i) \\ \alpha_i \\ \alpha_i/(1+\alpha_i) \end{pmatrix} \tag{4.17}$$

where the k-index has been omitted with the convention that $k = 1$ is the 1^{st} row of the above column vector, $k = 2$ is the second row, etc. Here τ_i and α_i are the time and phase angle of the i^{th} observation, f_j is the j^{th} frequency, etc. Likewise, the linear-parameter matrix X in this case is

$$X_j = \begin{pmatrix} H_j \\ (A_{1,1})_j \\ (A_{2,1})_j \\ (A_{1,2})_j \\ (A_{2,2})_j \\ \beta_j \\ C_j \end{pmatrix} \tag{4.18}$$

where H_j is the fitted absolute magnitude for the j^{th} frequency, etc.

The general LLSq solution to Equation (4.16) is

$$X_{jk} = \sum_{\ell,n,p} S_{jk\ell} L_{nj\ell} (B^{-1})_{np} m_p, \tag{4.19}$$

where B^{-1} is the inverse of the data-covariance matrix B:

$$B = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N_{\rm obs}}^2 \end{pmatrix}, \tag{4.20}$$

and $S_{jk\ell}$ is the parameter-covariance matrix, given by

$$S_{jk\ell} = [(s_j)^{-1}]_{k\ell}, \tag{4.21}$$

where in the above definition we invert each of the $N_{\rm frq}$ matrices s_j, these being defined by

$$(s_j)_{k\ell} \equiv \sum_{n,p} L_{njk}(B^{-1})_{np}L_{pj\ell}. \tag{4.22}$$

The elements of the parameter-covariance matrix S are the variances and covariances of the fitted parameters (as a function of frequency). The fit's residuals (as a function of frequency) are:

$$R_{ij} = m_i - \sum_k L_{ijk}X_{jk}, \tag{4.23}$$

and the fit's chi-squared (as a function of frequency) is:

$$(\chi^2)_j = \sum_{\ell,n} R_{\ell j}(B^{-1})_{\ell n}R_{nj}. \tag{4.24}$$

The frequency-dependent chi-squared $(\chi^2)_j$ is also known as the *periodogram*. Formally, the best-fit rotation frequency corresponds to the minimal value of $(\chi^2)_j$, but this may differ from the *preferred* frequency solution if the lightcurve is contaminated by other systematic periodic signals, if the data suffer from underestimated measurement uncertainties, or if the best-fit frequency corresponds to a dominant first harmonic (as opposed to a preferred dominant second harmonic, cf. Section 4.3.1.2).

Figure 4.3 details our iterative lightcurve-fitting algorithm's logic. Fitting com-

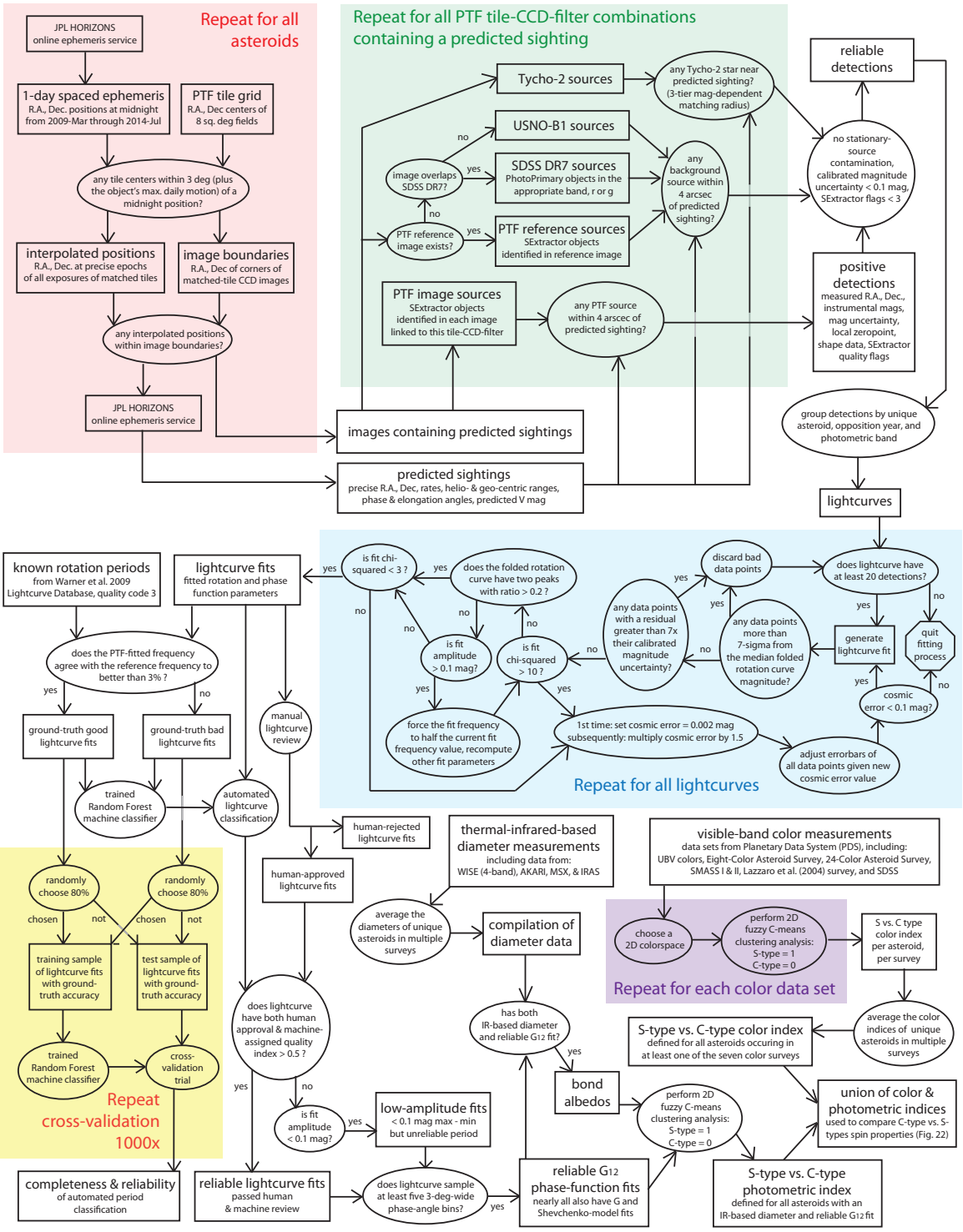


Figure 4.3: Diagram detailing the logic of this work's data reduction and analysis. Includes mining the survey for known-asteroid observations, aggregation of the data into lightcurves, vetting of the lightcurves and an application wherein phase functions are compared to color-derived asteroid taxonomy. See text for details.

mences as long as 20 or more 'reliable' data points (cf. Section 4.2.2 and Figure 4.2) are associated with a lightcurve. Irrevocably-bad data points are discarded in the first round of iterations, these include detections with 7σ or greater residuals from the initial solution. Examples of detections with such high residuals include contamination from background sources missing in the reference catalog, bad detector pixels that were not flagged by the pipeline, or spurious zeropoint solutions.

In the next stage of iterations, the fit's χ^2 per degree of freedom is reduced to $\sim$1 (formally, it is reduced until it is less than 3, cf. Figure 4.3) by gradually inflating the observations' errorbars through addition of a 'cosmic error', so-named because it encompasses contamination from possible errors (in all the 'cosmos'). In general the cosmic error represents the same diverse contaminating phenomena responsible for the $>7\sigma$ deviations seen in the initial iterations (cf. previous paragraph) just to a lesser extent.

Separately, this errorbar inflation compensates for our model's inability to fit each asteroid's precise periodic structure using only two harmonic terms in the Fourier series. In the limit of infinite observations and sufficiently many Fourier terms, we would ideally expect our data's errorbars to reflect true Gaussian variance. However, by truncating the series at two harmonics and using sufficiently precisely-calibrated photometry, we are in effect choosing to sacrifice (downsample) some of our photometric precision to obtain a formally better fit at the coarser resolution limit of the model.

To illustrate use of the cosmic error, consider the example of an eclipsing binary lightcurve, i.e., a rotation curve which is effectively sinusoidal *except* for a small interval around the phase of minimum flux, when it dips to a lower-than-predicted brightness. Examples from our dataset appear in Figure 4.10. Observations acquired during such eclipses will have systematic negative deviations greater in absolute value than would be explained by Gaussian variance alone. Increasing the errorbars of these observations will decrease the fits' χ^2 without altering the value of the fitted frequency. The fitted parameters' uncertainties (both for frequency and the linear parameters) are accordingly inflated as a penalty, and the fitted amplitude will be

underestimated. As detailed in Figure 4.3, the initial cosmic error used is 0.002 mag, and each iteration it is multiplied by a factor 1.5 until the χ^2 is sufficiently low. If the cosmic error exceeds 0.1 mag, the fitting is aborted. If the χ^2 (per degree of freedom) drops below 3 while the cosmic error is still below 0.1 mag, the fitting process concludes 'successfully' (see Figure 4.3).

Concurrently, each iteration includes a test for the presence of double peaks in the folded rotation curve (only if the fitted amplitude is at least 0.1 mag). In particular, if there exist two maxima and two minima in the folded lightcurve, we demand that the ratio of these peaks be greater than 0.2. Such a solution is preferred (cf. Section 4.3.1.2) given our ellipsoidal shape assumption, as described by [Harris et al. (2014)].

Denote as $f_{\rm best\_global}$ the frequency yielding the absolute minimum χ^2 per degree of freedom value, denoted $\chi^2_{\rm min\_global}$ (after the cosmic error has been tuned). If the folded lightcurve is single-peaked (or has only a relatively small secondary peak), then another deep minimum usually exists at the harmonic frequency $f_{\rm best\_harmonic} = 0.5 \times f_{\rm best\_global}$, the local minimum χ^2 value of which we denote $\chi^2_{\rm min\_harmonic}$). For cases wherein $\chi^2_{\rm min\_harmonic} < \chi^2_{\rm min\_global} + \text{inv-}\chi^2\text{-cdf}(0.95, 7)$, where inv-$\chi^2$-cdf$(p, N)$ is the inverse of the χ^2 cumulative distribution function for N free parameters evaluated at p, then we instead choose $f_{\rm best\_harmonic}$ rather than $f_{\rm best\_global}$. The 1σ uncertainty interval for the best-fit frequency is then found by computing the upper and lower intersections between $\chi^2_{\rm min} + \text{inv-}\chi^2\text{-cdf}(0.68, 7)$ and the periodogram in the vicinity of $f_{\rm best}$. Note that we used $N = 7$ free parameters in this case, i.e., the number of elements of X_j (Equation 4.18).

4.4.2 Nonlinear phase-function parameters

Modeling the phase function ϕ with either the G or G_{12} model (Equations [11] and [12]), introduces a second non-linear parameter (after the frequency f) and so we must modify the equations of the previous section accordingly. We sample $N_{\rm pha} = 200$ evenly-spaced phase-function parameter values. In particular, for G we test the interval $-0.3 \leq G \leq 0.7$ in steps of $\Delta G = 0.005$, and for G_{12} we test the interval

$0 \leq G_{12} \leq 1$ in steps of $\Delta G_{12} = 0.005$.

Our approach is to modify the left-hand side of Equation (4.16) by defining a new matrix m'_{iq} which contains all possible phase-function-corrected observed magnitudes:

$$m'_{iq} \equiv m_i - \Phi_{iq} = \sum_{j,k} L_{ijk} X_{jkq} \begin{cases} i = 1, 2, ..., N_{\rm obs} \\ j = 1, 2, ..., N_{\rm frq} \\ k = 1, 2, ..., N_{\rm lin} \\ q = 1, 2, ..., N_{\rm pha} \end{cases} \tag{4.25}$$

where, e.g., for the case of the G-model (Equation [4.11]),

$$\begin{aligned} \Phi_{iq} &\equiv -2.5 \log_{10}[\phi(\alpha_i, G_q)] \\ &= -2.5 \log_{10}[(1 - G_q)\phi_1(\alpha_i) + G_q \phi_2(\alpha_i)] \end{aligned} \tag{4.26}$$

The linear-parameter-solution array X now has an extra index q, reflecting the fact that we are now solving for each linear parameter as a function of the two non-linear parameters. The design matrix has the same number of indices as before (but fewer rows):

$$L_{ij} = \begin{pmatrix} 1 \\ \sin(2\pi f_j \tau_i) \\ \cos(2\pi f_j \tau_i) \\ \sin(4\pi f_j \tau_i) \\ \cos(4\pi f_j \tau_i) \end{pmatrix}, \tag{4.27}$$

while the linear-parameter matrix X is now

$$X_{jq} = \begin{pmatrix} H_{jq} \\ (A_{1,1})_{jq} \\ (A_{2,1})_{jq} \\ (A_{1,2})_{jq} \\ (A_{2,2})_{jq} \end{pmatrix}. \tag{4.28}$$

The appeal in adopting the above approach is that the general solution is only

slightly modified:

$$X_{jkq} = \sum_{\ell,n,p} S_{jk\ell} L_{nj\ell} (B^{-1})_{np} m'_{pq}, \tag{4.29}$$

where the only difference between equations (19) and (29) are the q indices appended to X and m (and the latter being redefined as m').

The fit's residuals R are now a function of frequency *and* phase-function parameter:

$$R_{ijq} = m'_{iq} - \sum_{k} L_{ijk} X_{jkq}, \tag{4.30}$$

as is the fit's chi-squared:

$$(\chi^2)_{jq} = \sum_{\ell,n} R_{\ell jq} (B^{-1})_{\ell n} R_{njq}. \tag{4.31}$$

As a function of any of the *linear* parameters, the fit's χ^2 varies precisely quadratically, whereas as a function of frequency it has an intricate spectral structure with many local minima. As a function of a non-linear phase parameter (G or G_{12}), the χ^2 tends to have a single minimum (on the range we evaluate): in this sense G and G_{12} are more similar to the linear parameters than they are to frequency. However, the generally asymmetric shape of the phase parameter' χ^2 dependence necessitates its grid-based numerical treatment—particularly to ensure accurate estimation of the phase parameter's uncertainty.

The two-dimensional χ^2 surface given by Equation (4.31), which is defined on a $N_{\text{freq}} \times N_{\text{pha}}$ grid, can be reduced to a one-dimensional χ^2 function by choosing, for each frequency index j, the phase-parameter index q that minimizes the χ^2. The result is a one-dimensional periodogram, as in Equation (4.24). Once the fitted frequency is identified, we compute the uncertainty in the fitted f by the method described in the previous section using the inv-χ^2-cdf() function. We then likewise numerically compute the uncertainty in the phase parameter by again collapsing $(\chi^2)_{jq}$ to a one-dimensional vector, this time as a function of the phase parameter with the frequency

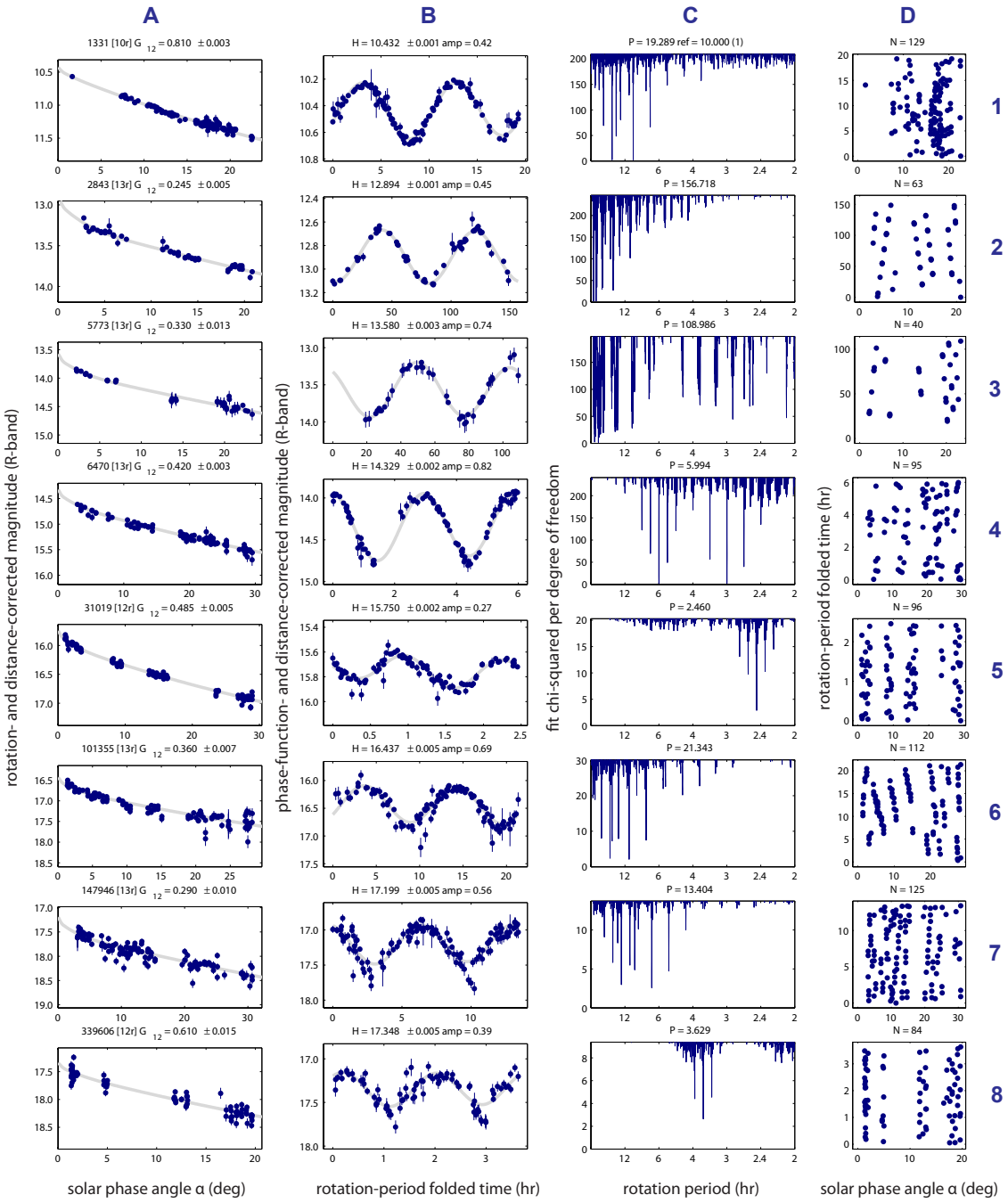


Figure 4.4: Examples of lightcurves having both well-sampled rotation *and* phase-function components. Each row corresponds to a different asteroid. These example asteroids are sorted vertically by their physical diameter (assuming 7% albedo); the top object is ∼45 km and the bottom object is ∼2 km. Column A shows the phase curve (corrected for rotation); Column B shows the rotation curve (corrected for phase-function); Column C shows the periodogram; Column D shows the distribution of the observations in rotational phase vs. solar phase angle. Above each plot is additional information depending on the column: (A) the asteroid number, followed by (in square brackets) the opposition year (most are 2013) and filter (in all cases 'r') followed by the fitted G_{12} parameter; (B) the fitted absolute magnitude and amplitude; (C) the fitted period (in hours); (D) the number of data points included (and shown) in the fit.

fixed at the fitted value (j-index), and use the inv-χ^2-cdf() function to estimate the uncertainty in the phase parameter.

As noted in Table 4.1, a total of 587,466 lightcurves exist in PTF, where each lightcurve by definition consists of all reliable observations of a unique asteroid observed in a single opposition in a single photometric band. Of these, only ∼10% (59,072 lightcurves) have at least 20 observations and therefore qualified for fitting with our algorithm. A total of 54,296 lightcurves actually produced a fit—the remaining ∼5,000 lightcurves failed to produce a fit either because some observations were discarded and the total fell below 20 data points, or because the fitted cosmic error grew to exceed 0.1 mag.

Figure 4.4 shows several examples of lightcurves fitted with the algorithm described in this section. In the third column (column C) of Figure 4.4, we show the periodograms of each lightcurve. Note that although the periodogram's horizontal axes are labeled with the *period* (for easier interpretation), the chi-squared (per degree of freedom) values are actually plotted linearly with respect to *frequency*. This is because, as described earlier, our sampling is uniform with respect to frequency, and the harmonics are more easily seen with constant frequency spacing. Column (D) shows the data sampling in rotational phase versus solar phase angle, a useful plot to ensure there is no obvious correlation between the two (which could lead to an erroneous fit, e.g., for long periods, large amplitudes and/or few data points).

4.4.3 Comments on implementation

Each iteration in the fitting of each asteroid lightcurve involves evaluating the arrays and tensor-products in either Equation (4.19) or (4.29). This includes inverting the data-covariance matrix B (Equation [4.20]) and inverting the $N_{\rm frq}$ matrices s_j (Equation [4.22]). The arrays L, m', X and R can have a relatively large number of elements, making them and their relevant products potentially taxing with respect to computational memory.

Our particular implementation of this algorithm leverages the efficient array-

manipulation capabilities of MATLAB, especially its ability to perform fast matrix multiplication and matrix inversion utilizing BLAS calls[11] and OpenMP multi-threaded C loop code[12]. Given typical numbers of observations and frequency sampling, each of our lightcurve fits (including the multiple iterations) takes on average several tens of seconds to run on an eight-core machine (multi-threading enabled), and typically consumes less than ∼4 GB of memory using single-precision computation.

In the online supplementary material we provide our custom MATLAB function used for fitting the G-parameter version of the lightcurve model (`asteroid_lc_fit_G.m`). Analogous versions exist for the Shevchenko and G_{12} models. This function takes as input an asteroid's apparent magnitudes, magnitude uncertainties, observed epochs, phase angles, geocentric and heliocentric distances. Its outputs include the linear-parameter-solution array (Equation 4.28), residuals (Equation 4.30), chi-squared array (Equation 4.31), and additional information about each lightcurve solution such as the amplitude and peak ratios.

4.5 Reliability of fitted rotation periods

A primary concern in the quality assessment of our fitted lightcurve parameters is the validity of our derived rotation periods. In this section we describe several methods of estimating the reliability of these periods, beginning with comparison to a ground-truth subsample of known-period asteroids and followed by a full vetting of our entire sample using a combination of machine-learning and manual classification.

The fitted period may differ (slightly or significantly) between the fits using the different phase function models. In this section for simplicity we consider only the period value obtained when fitting with the G_{12} phase-function model (Section 4.3.2.3). In subsequent sections we will again consider all three ϕ models.

[11] http://www.netlib.org/blashttp://www.netlib.org/blas
[12] http://openmp.orghttp://openmp.org

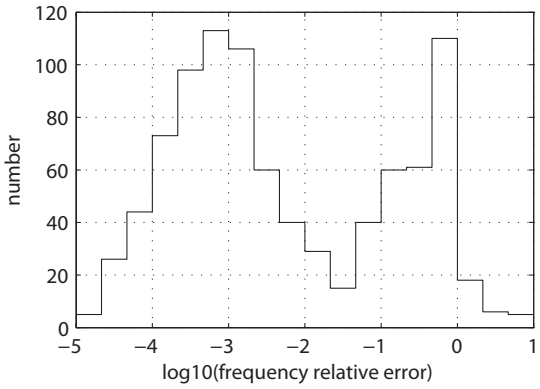


Figure 4.5: For the 927 lightcurves (805 unique asteroids) having a quality code 3 period in the Lightcurve Database of [Warner et al. (2009)] *and* an original fit in this work, we plot the distribution of the relative error in our fitted rotation frequencies with respect to the literature-referenced frequencies. The distribution is bimodal, with the left-hand mode corresponding to those fits having better than ∼3% agreement.

4.5.1 Known-period subsample

A total of 927 (∼2%) of our fitted lightcurves belong to 805 unique asteroids having a previously-measured period listed in the Lightcurve Database (LCBD) of [Warner et al. (2009)]. This includes only asteroids having a quality code of 3 (highest quality) in the LCDB.

Figure 4.5 shows that the distribution of relative errors on our fitted frequencies is bimodal, with the left mode corresponding to periods having better than ∼3% agreement with the reference period, and the right mode corresponding to periods in disagreement with the reference period. These disagreeing fits include lightcurves which differ from the reference value by a harmonic (half = relative error 0.5, double = relative error 1.0), as well as frequencies that do not differ by a factor of two or any integer multiple. About 1/3 of the lightcurves in Figure 4.5 fall into the right mode and are thus considered disagreeing fits.

Figure 4.6 shows some examples of these disagreeing fits. Row 1 shows an apparent low-amplitude rotator, whose fitted period of 15.7 hr differs from the reference value of 9.7 hr. Row 2 is an object whose periodogram contains a great deal of noise, divided into two broad forests of frequency minima. The left forest appears to have

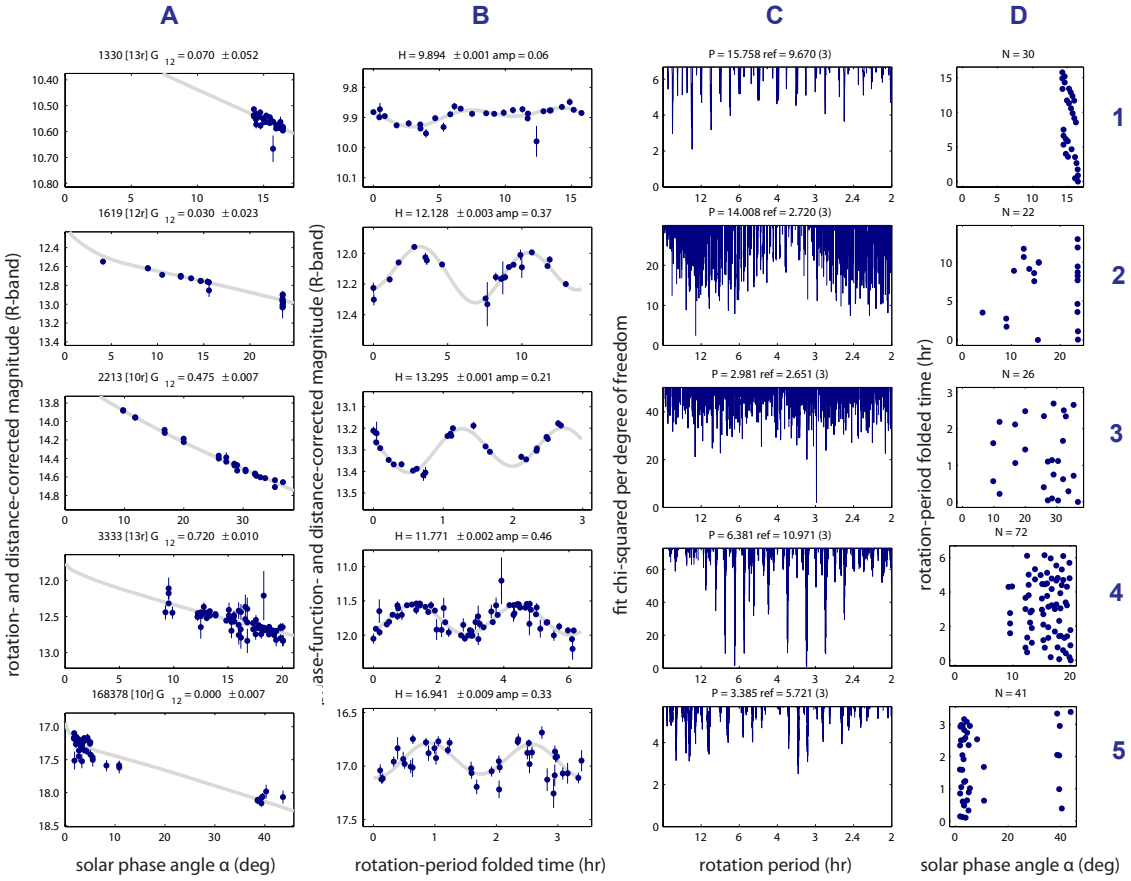


Figure 4.6: Examples of lightcurves whose fitted frequency differs from the reference frequency by more than 3%, so that they fall in the right mode in the histogram shown in Figure 4.5 and are formally defined as inaccurate fits. *Row 1*: Low-amplitude rotator. *Row 2*: Incorrect period (too few observations?). *Row 3*: A fitted frequency that differs from the reference frequency by 12%. *Row 4*: period that differs by a non-integer multiple, despite looking reasonable. *Row 5*: Folded lightcurve appears to be fitting noise in the data.

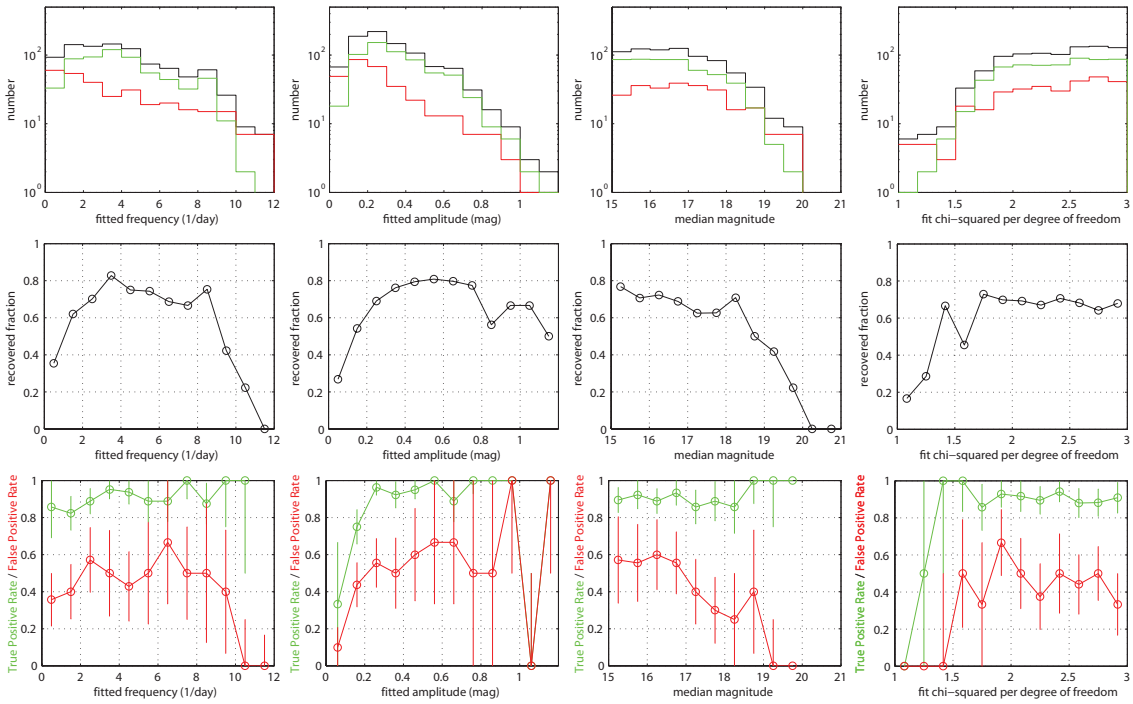


Figure 4.7: *Top row*: The 927-lightcurve known-period sample (black), divided into the accurately-fitted (green) and inaccurately-fitted (red) subgroups. *Middle row*: Ratio of the green to black histograms. *Bottom row*: Results of cross-validation of the machine-classifier (see Section 4.5.2.2).

been selected by our fitting algorithm while the right forest seems associated with the true period of ∼2.7 hr. Row 3 contains an object whose 12% relative frequency error exceeds the 3%-accuracy threshold we have defined, and so despite appearing to be a good fit it is formally categorized as inaccurate. Row 4 also looks like a reasonable fit at 6.4 hr, but disagrees with the reference period of 11.0 hr (though the latter does have a perceptible local minimum in the periodogram). Finally, Row 5 includes a likely example of the algorithm fitting noise in the photometry of a faint asteroid.

In Figures 7 and 8 (top and middle rows) we detail the distribution of the accurately-recovered-period and inaccurately-recovered-period subgroups in terms of eight different lightcurve parameters. Some basic observations from these histograms are:

1. fitted periods are far less reliable if longer than ∼1 day or shorter than ∼2.7 hours,

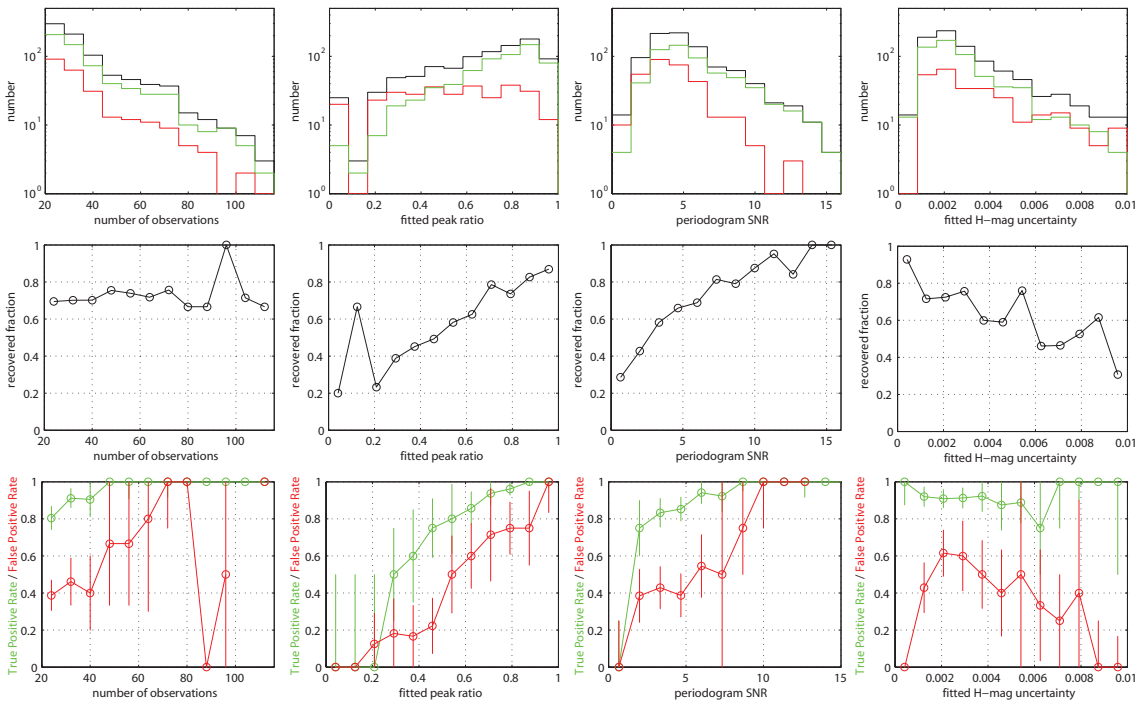


Figure 4.8: *Top row*: The 927-lightcurve known-period sample (black), divided into the accurately-fitted (green) and inaccurately-fitted (red) subgroups. *Middle row*: Ratio of the green to black histograms. *Bottom row*: Results of cross-validation of the machine-classifier (see Section 4.5.2.2).

2. fitted amplitudes of less than 0.1 mag correspond to the least reliably fit periods,

3. lightcurves consisting of observations dimmer than ∼18.5 mag are much less reliable than brighter lightcurves (though they are also far less numerous in the known-period sample),

4. fit χ^2 (per degree of freedom) values of less than ∼1.7 correlate with less reliable periods (though they are also far less numerous in the known-period sample). Note that, in the fitting process, growth of the cosmic error term ceased once the χ^2 (per degree of freedom) fell below 3 (cf. Figure 4.3).

5. the number of observations in a lightcurve is *not* directly correlated to the reliability of the fitted period,

6. the ratio of the folded lightcurve's two peaks, the signal-to-noise ratio of the periodogram's chosen minimum, and the uncertainty in the absolute magnitude parameter are all strong indicators of the reliability of the fitted period.

The above comments reflect consideration of the *one-dimensional* distributions in Figure 4.7 and 8; however we can easily imagine there are correlations in more dimensions not evident from these plots alone. An obvious example would be the two-dimensional distribution in amplitude versus median magnitude: reliability is presumably greater for bright asteroids having amplitudes <0.1 mag than it is for dim asteroids having amplitudes <0.1 mag. Period versus amplitude is also likely an insightful distribution (and was considered for example by [Masiero et al. (2009)]). The number of observations possibly *does* correlate with reliability if we were to restrict another parameter or parameters to some specific interval.

Rather than manually examining the period-fitting reliability as a function of all possible multi-dimensional combinations of the eight lightcurve parameters detailed in Figures 7 and 8, we can take a more general approach of considering the reliability to be a single function defined on the multi-dimensional parameter space in which all of the lightcurves reside. We hypothesize that accurately-fit lightcurves and inaccurately-fit lightcurves occupy distinct regions in this multi-dimensional volume.

As these volumes can overlap to some extent, we can at least estimate the *probability* that a lightcurve with that particular vector of parameters corresponds to an accurately-recovered (or inaccurately-recovered) period when obtained by the fitting algorithm of Section 4.4.

There are two general ways of accomplishing this goal. One way is to produce a large number of synthetic lightcurves filling out the multidimensional lightcurve-parameter space, subject these synthetic lightcurves to our fitting algorithm, and thereby map out e.g., by binning and interpolation, the fit reliability throughout the multi-dimensional volume. This method requires us to accurately simulate all sorts of varying sampling cadence as well as measurement uncertainties, including contributions from both systematics and noise, and it requires significant extra computing time to actually subject the synthetic data to our fitting procedure. The second method—the approach we take in this work—uses a ground-truth sample (the known-period lightcurves already described in this section) to train a machine classifier to discriminate reliable versus unreliable fits within the multi-dimensional lightcurve-parameter space.

4.5.2 Machine learning

We adopt a supervised ensemble-method approach for classification, originally popularized by [Breiman et al. (1984)], specifically the *random forest* (RF) method [Breiman (2001)]. RF classification has extensive and diverse applications in many fields (e.g., economics, bioinformatics, sociology). Within astronomy in particular RF classification is one of the more widely-employed methods of machine-learning, though many alternatives exist. For example, [Masci et al. (2014)] use the RF method for variable-star lightcurve classification, while others have approached this problem via the use of, e.g., support vector machines [Woźniak et al. (2004)], Kohonen self-organizing maps [Brett et al. (2004)], Bayesian networks and mixture-models [Mahabal et al. (2008)], principle component analysis [Deb & Singh (2009)], multivariate Bayesian and Gaussian mixture models [Blomme et al. (2011)], and thick-pen

Table 4.2: Summary of the 20 lightcurve parameters (features) used by our period-quality classifier. See text for a discussion of the cross-validation-derived importance value (Section 4.5.2.2).

feature	importance (%)	description
`peakRatio`	11.1	Ratio of the fitted lightcurve's two peaks (= max − min). Zero if only one peak, one if exactly the same height.
`amplitude`	10.2	Fitted amplitude of the folded lightcurve. Equivalent to the height (max − min) of the larger of the two peaks.
`periodFit`	8.6	Rotation period value obtained using this work's data and fitting algorithm.
`freqSNR`	8.4	Signal-to-noise of the fitted (minimum) frequency in periodogram = $2 \times \lvert\text{min} - \text{median}\rvert/(84^{\text{th}}\text{-percentile} - 16^{\text{th}}\text{-percentile})$
`hMagErr`	5.8	Uncertainty in the fitted H-magnitude (i.e., error in the fitted absolute magnitude)
`a12Coeff`	4.3	Fourier coefficient A_{12}
`a22Coeff`	4.2	Fourier coefficient A_{22}
`numObsFit`	4.1	Number of observations in the final fitted lightcurve, after discarding any bad observations
`medMag`	4.1	Median calibrated magnitude (in the photometric band specific to the lightcurve, either R or g)
`chisq`	4.1	Reduced chi-squared of the fit i.e., χ^2 per degree of freedom)
`a21Coeff`	4.0	Fourier coefficient A_{21}
`a11Coeff`	3.8	Fourier coefficient A_{11}
`rmsFit`	3.8	Root-mean-squared residual of the fit
`hMagRef`	3.8	Reference H-magnitude (i.e., absolute magnitude of the asteroid in V-band as listed by the MPC)
`kIndex`	3.7	Stetson's K-index (a measure of kurtosis in the magnitude distribution of a folded lightcurve, introduced by [Stetson (1996)].)
`freqResol`	3.7	Resolution of the periodogram: $\Delta f = 1/(4\Delta t)$ where Δt is the time between the first and last observations in the lightcurve
`hMagResid`	3.7	Difference between the reference absolute magnitude (`hMagRef`) and the fitted H-magnitude
`cuspIndex`	3.6	'Cusp index': Median squared residual of the dimmest 10% points divided by the median squared residual of all other points
`numObsRem`	2.9	Number of observations removed during the fitting process (due to >7-sigma residuals with respect to preliminary fits)
`cosmicErr`	2.1	Final 'cosmic error' value at end of fitting process (<0.1 mag in all cases)

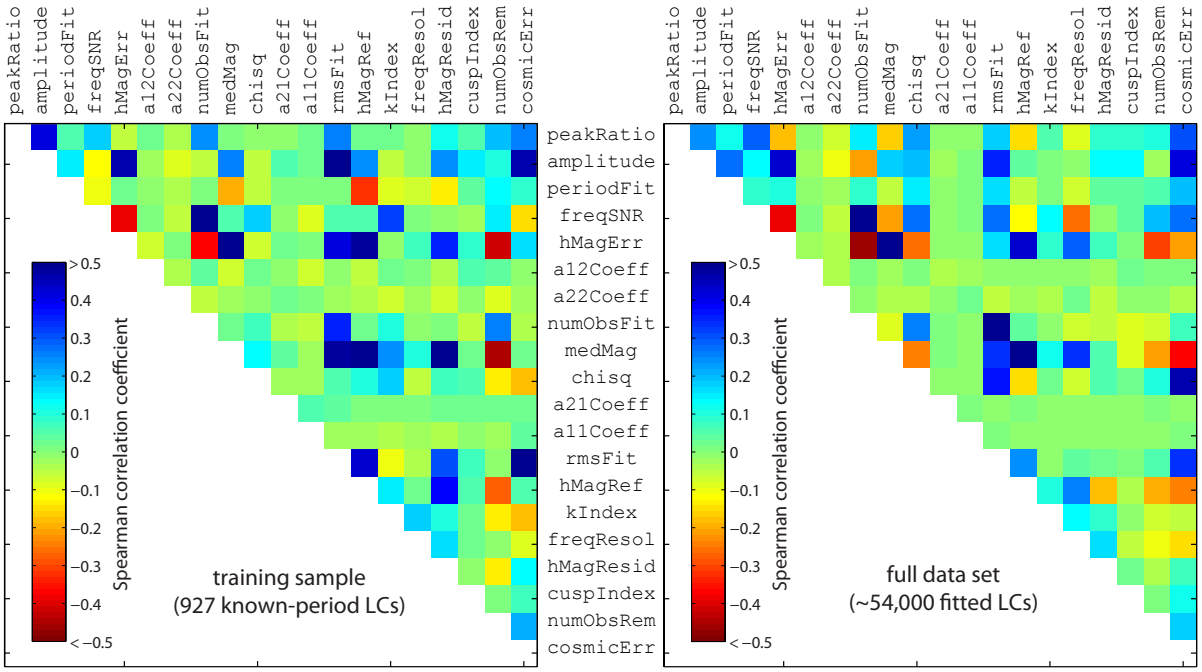


Figure 4.9: Correlation matrices (Spearman's ρ coefficient) for the 20 lightcurve features (Table 4.2) in the training sample (left) and in the full data set (right).

transform methods [Park et al. (2013)].

For general descriptions of RF training and classification, we refer the reader to [Breiman (2001)], [Breiman & Cutler (2004)], and the many references cited by [Masci et al. (2014)]. Our use of a RF classifier is particularly motivated by its already-proven application to the discovery and classification of astrophysical transients in the same PTF survey data [Bloom et al. (2012)], as well as streaking near-Earth asteroid discovery in PTF data (Waszczak et al. in prep.).

Machine-learning application generally consists of three stages: training, cross-validation, and classification. In the training stage of building a machine classifier, the multi-dimensional parameter space is hierarchically divided into subspaces called *nodes*, these nodes collectively comprise a *decision tree*. The smallest node—also known as a *leaf*—is simply an individual datapoint (in our case, a single lightcurve). Given a set of leaves with class labels, one can build an ensemble of trees (called a *forest*), each tree representing a unique partitioning of the feature space, wherein the

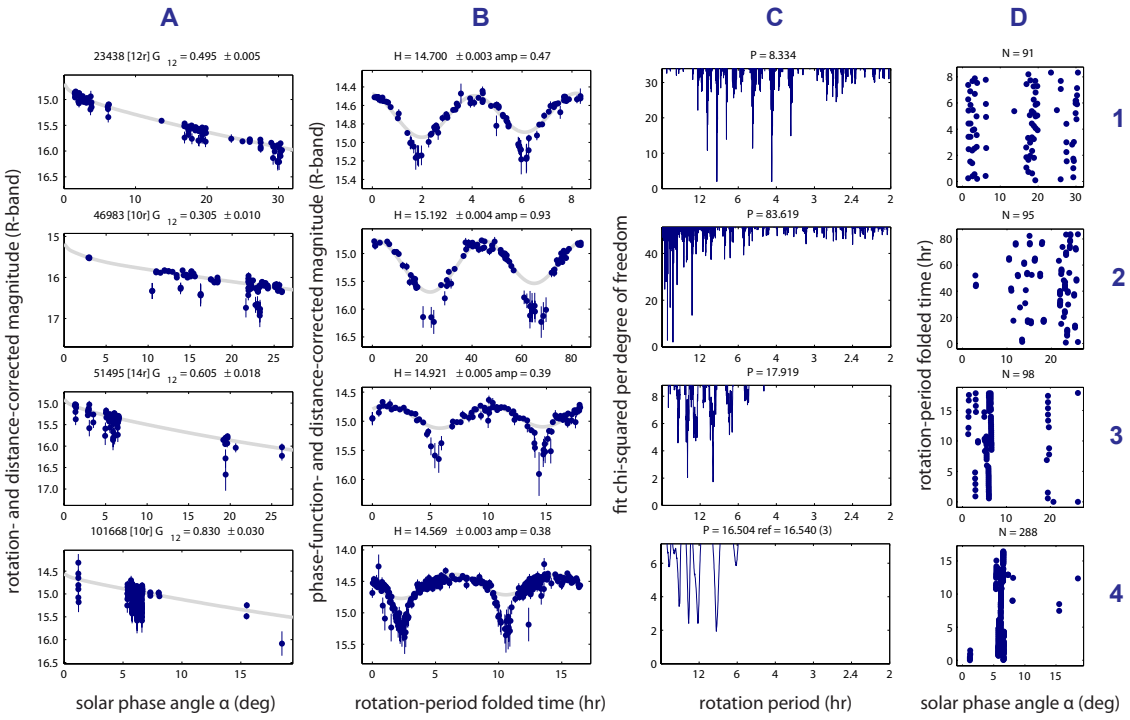


Figure 4.10: Examples of reliable lightcurves whose folded rotation curve include cusp-like minima (systematic negative deviation from the 2nd-order Fourier fit at minimum brightness), suggestive of a binary system. Many more examples exist in our lightcurves, however in this work we have not specifically flagged such lightcurves. Future works will more carefully label and analyze this particular class of objects.

nodes are split with respect to different randomly-chosen subsets of the parameter list. Each node splitting attempts to maximize the separation of classes between the sub-nodes. Serving as a model, in the subsequent classification stage the forest allows one to assign a probability that a given vector of features belongs to a given class. During cross validation (an essential early stage in this process), the training and classification steps are repeated many times, each time using different subsamples (of labeled data) as the training data and testing data. Cross validation evaluates the classifier's performance and ensures it is not overfitting the training data.

For our lightcurves, we are interested in a binary classification, i.e., whether the fitted period is accurate ('real') or inaccurate ('bogus'). [Bloom et al. (2012)] coined the term `realBogus` to describe this binary classification probability in the context of extragalactic transient identification. In the present work we are essentially adapting Bloom et al.'s `realBogus` concept to the problem of lightcurve-period reliability assessment.

We employ a MATLAB-based Random Forest classifier[13] which is a port of the original RF software (originally implemented in R). This software includes two main functions, which perform the training and classification steps separately.

4.5.2.1 Classifier training

Our training data consist of the known-period lightcurves (cf. the previous section) belonging to the two classes under consideration: 618 lightcurves having accurately-fit rotation periods and 309 lightcurves having inaccurately-fit periods. Membership in one class versus the other depends on our arguably arbitrary 3% relative accuracy threshold, though we claim the clearly bimodal shape of the distribution in Figure 4.5 justifies this 3% criterion. We note also that the classifier ultimately only provides a *probability* that a given lightcurve belongs to one class or the other, so that objects very near to the 3% cutoff may conceivably correspond to classification probabilities close to 0.5.

[13] https://code.google.com/p/randomforest-matlabhttps://code.google.com/p/randomforest-matlab

An important point is that the 'ground-truth' reference periods we have taken from the database of [Warner et al. (2009)] may include some number of inaccurate periods. Such periods may be the product of erroneous fitting on the part of any one of its many different contributors, each of whom may employ a different fitting procedure and/or adhere to different confidence criteria. For the sake of this work however we consider all quality code 3 periods to be accurate—any actual inaccuracy will contribute to decreased classifier performance.

Besides ground-truth periods that are simply inaccurate, we also in principle risk contamination from reference periods that are *no-longer accurate*. We assume that the majority of asteroids' periods are not changing with time, at least not at levels measureable with our data. For instance, direct measurement of the YORP mechanism in at least one asteroid [Lowry et al. (2007)] reveal a relative rotation period change of several parts per million over several years. Any *measureable* period changes would likely be due to recent collisional events. The case of asteroid 596 Scheila [Bodewitts et al. (2011)] demonstrates that detectable collisional events among main-belt asteroids do occur on a relatively regular basis, though even this robustly-detected collision imparted no measurable change in the asteroid's spin rate [Shevchenko et al. (2013)].

Although Figures 7 and 8 detail the period-fitting reliability as a function of only eight lightcurve parameters, we construct our classifier using 12 additional parameters, for a total of twenty lightcurve parameters. In the context of machine-learning these parameters are known as *features*. The twenty features we use were chosen on the basis of their availability (most are output directly by the fitting process and do not require additional computation) as well as their actual importance (as computed during the cross-validation tests described in the next section).

Our twenty lightcurve features are listed in Table 4.2, in order of decreasing importance. Most of these quantities we have discussed already in previous sections in the context of our model and fitting procedure. The list also includes two features characterizing the magnitude distribution of the folded lightcurve: (1) Stetson's K-index, a measure of the kurtosis borrowed from variable star lightcurve analy-

		classified (predicted) as:	
		accurate period	inaccurate period
actual accuracy (ground truth):	accurate period	true positive (TP)	false negative (FN)
	inaccurate period	false positive (FP)	true negative (TN)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

Figure 4.11: Definitions of true vs. false and positive vs. negative labels. True-positive rate (TPR) is sometimes called the *completeness* or *sensitivity*, while false-positive rate (FPR) is otherwise known as the *false-alarm rate*, one minus the *reliability*, or one minus the *specificity*.

sis [Stetson (1996)], and (2) a 'cusp index' which quantifies the extent to which the dimmest 10% of the data points in the folded lightcurve deviate from the best fit relative to the other 90% of the data points. We designed the cusp index to potentially identify eclipsing systems which are poorly fit by the two-term Fourier approximation but nonetheless may have accurately-fit periods (examples of lightcurves with such cusp-like minima appear in Figure 4.10). Eclipsing binaries would be most properly treated with a different model entirely, as would tumbling asteroids (which we also did not systematically try to identify in the data, and probably lack reliable lightcurve solutions when subjected to this work's algorithm).

Figure 4.9 visualizes the two-dimensional correlation coefficients for all possible pairs of the 20 lightcurve features. Overall, the correlation structure of the training sample qualitatively resembles that of the full data set, implying the training set fairly well represents the overall data set in terms of its feature-space structure. On the other hand, the *distributions* (e.g., median value, range of values) of individual features in the training set do not necessarily match the distributions in the full data set: this is evident for the several features plotted in Figure 4.14. An obvious example is that the full data set contains far more faint asteroids than does the training sample, even though in both cases the median magnitude (`medMag`) is positively correlated with quantities like `rmsFit` (due to Poisson noise) and `hMagRef` (since larger asteroids

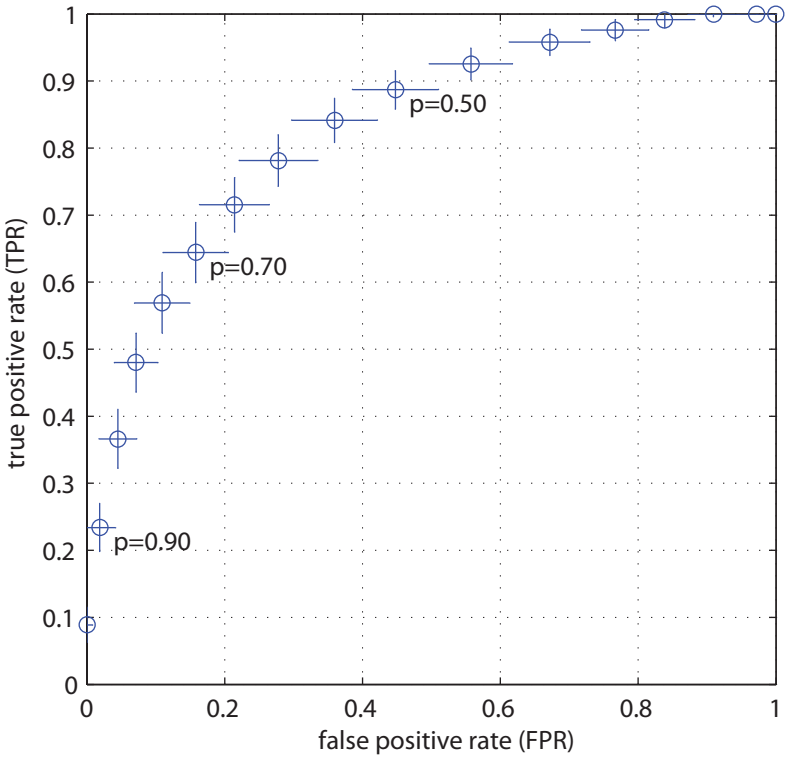


Figure 4.12: True-positive versus false-positive rates for the cross-validation trials. Such as plot is sometimes referred to as a *receiver operating characteristic* (*ROC*) curve. Each trial trains the classifier using a randomly-chosen 80% of the known accurate fits and 80% of the known inaccurate fits among the 927-lightcurves that have reference periods. The 20% remaining lightcurves serve as the test sample. Moving along the hyperbolic locus of points in this plot is equivalent to tuning the classification probability threshold from zero (lower left of the plot) to one (upper right of the plot). The errorbars represent the scatter in the 1,000 cross-validation trials.

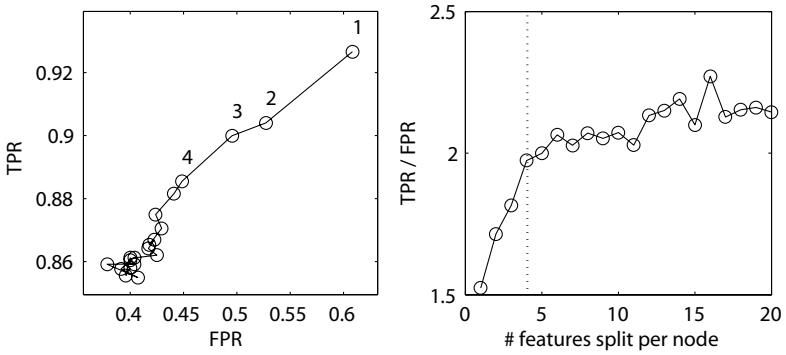


Figure 4.13: Varying the number of features that are randomly split per node in the decision-tree-building process affects both the TPR and FPR. The values plotted here correspond to the $p > 0.5$ classification threshold; each point was generated by the exact same process for which the results in Figure 4.12 were generated, only varying the number of features with respect to which nodes are split. In the left plot, the first four points are labeled with the number of features for that trial (for $n > 4$ we omit the label). In our actual implemented model (Figure 4.12) we chose $n = 4$ features, the value after which the TPR/FPR ratio plateaus at approximately 2, and also the value [Breiman (2001)] recommends, i.e., the square-root of the total number of features (in our case, $\sqrt{20} \approx 4$).

tend to be brighter).

4.5.2.2 Classifier cross-validation

To ascertain the trained classifier's capabilities, and to ensure that the classifier is not overfitting the training data, we perform a series of 1,000 cross-validation trials. In each trial we split each class (accurate fits and inaccurate fits) into a *training* subsample (a randomly chosen[14] 80% of the class) and a *test* subsample (the remaining 20% of the class). We then train a classifier using the combined training subsamples and subsequently employ the classifier on the combined test subsamples. In each of the trials, the classifier outputs a classification probability (score) for each object in the test sample, and we track the true positive rate (TPR; fraction of accurate period fits that are correctly classified above some threshold probability) as a function of the false-positive rate (FPR; fraction of inaccurate period fits that are incorrectly

[14]Another standard, slightly different approach is to evenly split the training data into k disjoint sets (a process called k-folding). Also, our choice to *separately* partition the two classes into training and test subsamples could be omitted.

classified above said threshold probability). See Figure 4.11 for a summary of these terms.

The results of the cross-validation are shown in Figure 4.12. By tuning the minimum classification probability used to threshold the classifier's output, one effectively moves along the hyperbola-shaped locus of points in TPR-vs.-FPR space seen in the plot. Several points have labels ($p = ...$) indicating the corresponding threshold probability (adjacent points being separated by $\Delta p = 0.05$). The errorbars in Figure 4.12 represent the standard deviation of the location of each point over all 1,000 trials, while the point centers are the average locations.

A classification threshold of $p > 0.5$ is conventionally used when quoting single false-positive and true positive rates. In our case, this gives FPR $= 0.45 \pm 0.07$ with TPR $= 0.89 \pm 0.03$. The *contamination* of positively-classified lightcurves in the cross-validation trials depends also on the actual class ratios in the sample being classified. In particular, since $\sim 1/3$ of our known-period lightcurves are inaccurate fits (Figure 4.5), it follows that among all lightcurves the classifier labels as accurate fits, the contaminated fraction is $(0.45 \times 1/3)/(0.89 \times 2/3 + 0.45 \times 1/3) \approx 1/5$. If instead of using the classifier we just randomly labeled some fraction of the lightcurves as accurate and the rest as inaccurate, the resulting contamination would be 1/3 (i.e., worse than the 1/5 afforded by the classifier, as expected).

Several parameters can be adjusted or tuned when training a random forest classifier. First is the number of decision trees generated during the training stage. Classification accuracy typically increases with the number of trees and eventually plateaus. Most applications employ hundreds to thousands of trees; we here use 1,000 trees. Another tunable parameter is the number of randomly-selected features (out of the 20 total here considered) with respect to which nodes are split in building the decision trees. [Breiman (2001)] recommends using the square root of the number of features. We ran the cross-validation for all possible numbers of features with respect to which the nodes can be split (i.e., all numbers between 1 and 20). The results are in Figure 4.13. We chose $n = 4$ as the number of features to split, both because the classifier's performance plateaus after that value and because it follow's

the recommendation of [Breiman (2001)] ($4 \approx \sqrt{20}$) features.

Other parameters that can be tweaked are the maximum depth of a tree, the minimum number of samples per leaf, the minimum number of samples used in a split, and the maximum number of leaf nodes. We do not constrain any of these parameters, meaning we allow: trees of any depth, with any number of leaf nodes, leaf nodes consisting of a single sample, and splits based on the minimum of 2 samples. We note that as a result our model optimization is not comprehensive and it is possible a better classifier could be trained. However, the relatively small training sample size here is likely the limiting factor; additional data is necessary to substantially improve the classifier performance.

In the bottom rows of Figure 4.7 and 4.8, we detail the dependence of the TPR and FPR on various lightcurve parameters. Averaging (marginalizing) over any of the x-axis quantities in these bottom-row plots (while also weighting each bin by the number of lightcurves it contains, cf. the top row of plots in Figures 7 and 8), produces precisely the TPR and FPR values of the $p = 0.5$ data point in Figure 4.12.

In addition to the TPR and FPR estimates, cross-validation allows us to quantify the relative *importance* of the features by computing the average depth in the trees at which a split was performed with respect to each feature. Those features with respect to which the training sample is consistently divided early in the building of each tree are deemed more important (i.e., more discriminating) than those features which are split later, as the tree-building process tries to maximize the separation of the classes as early as possible by splitting features in an optimal sequence. Both Table 4.2 and Figure 4.9 list the features in order of importance.

Note that we had manually guessed several of the most important features—namely, `peakRatio`, `freqSNR` and `hMagErr`—prior to any machine-learning work via inspection of the plots in Figure 4.8. The numerical importance values thus agree with these initial observations, and also quantify the significance of features which would be difficult to ascertain manually. For instance, `numObsFit` appears (in Figure 4.8) not to be related to the fitting accuracy while `medMag` (Figure 4.7) *does* appear related to accuracy (fainter lightcurves being less accurate), yet these two features

evidently have equal importance in the classification process (cf. Table 4.2). Figure 4.9 indicates that `numObsFit` and `medMag` have quite different correlation relationships with respect to more important features. Hence, it would not be surprising if their one-dimensional distributions (in Figures 7 and 8) bear no resemblance to the multi-dimensional distributions on which the decision trees are defined and in which these two parameters apparently carry comparable weight.

4.5.2.3 Machine-vetted lightcurves

Having trained the machine classifier as described in Section 4.5.2.1, we use it to predict the validity of our remaining ∼53,000 fitted periods (of ∼48,000 unique asteroids) which lack quality code 3 reference periods in [Warner et al. (2009)]. The automated classifier assigned positive reliability scores ($p \geq 0.5$) to 19,112 of the lightcurves (35% of the total data set). Figure 4.14 details the distribution of the lightcurves (raw-fitted, machine-vetted, and other subsets) with respect to some of the most important lightcurve features.

With respect to rotation period (Figure 4.14 panel A), the classifier rejects the largest fractions of lightcurves in the long-period ($\gtrsim$1 day) and short-period ($\lesssim$2.7 hours) bins. From Figure 4.7 (bottom row, leftmost column), we know that the classifier's completeness does not drop significantly for these long- and short-period objects, nor is the false positive rate higher among them. Hence we have reason to trust the classifier's heavy rejection of periods in these bins, and therefore conclude that our fitting algorithm (Section 4.4) is prone to erroneously fitting periods in these period extremes (as was also suggested in the known period sample in Figure 4.7).

Panel C shows that the mode of the apparent-magnitude (`medMag`) distribution for machine-approved lightcurves is ∼19 mag, as compared to the predominantly $V \lesssim 17$ mag known-period training sample. Comparing this to Figure 4.2 panel A shows that the limiting magnitude of reliable lightcurves is comparable to that of individual detections.

Panel E of Figure 4.14 shows that the raw output of our fitting process contains peak-ratio values that are uniformly-distributed above 0.2, this particular value being

a hard-coded threshold that double-peaked lightcurves (at least those with amplitudes >0.1 mag) output by our fitting algorithm must satisfy (see Figure 4.3 and Section 4.4.1). The classifier's output clearly indicates that reliability is linearly related to the peak ratio, as was also prominently seen in Figure 4.8. Because Figure 4.8 also indicates that the classifier's true-positive and false-positive rates also relate linearly with `peakRatio`, we conclude that the slope of the `peakRatio` distribution for the machine-vetted lightcurves is likely an upper limit for the true slope.

4.5.3 Manual screening

In addition to machine-based vetting, we manually inspected all 54,296 of the lightcurves that were output by our fitting process. A human screener first studies the ground-truth known period examples (Section 4.5.1) in an effort to learn to distinguish between accurate and inaccurate fits. Only the G_{12} fit is considered (as was the case with the automated classifier), and for each lightcurve the screener inspects precisely the amount of information included for example in Figures 4.4, 4.6 and 4.10 of this chapter. Specifically, for each lightcurve the screener views a row of four plots: (1) the rotation-corrected phase curve, (2) the phase-function-corrected folded rotation curve, (3) the periodogram, i.e., the reduced χ^2 plotted linearly against frequency (labeled however with the corresponding period), and (4) the rotational-phase vs. phase-angle plot. A single screener is presented with these plots through a plain-formatted webpage, allowing for efficient scrolling through the lightcurves and rapid recording of either a 'reliable' or 'unreliable' rating for each fitted period. In addition, all lightcurves in the known-period sample were reinserted into the screening list, with their reference periods removed. These were thus blindly assessed by the screener, independent of their formal (3%-accuracy) classification status.

The black lines in Figure 4.14 plot the results of the manual screening, in which a total of 10,059 lightcurves (19% of the total set) were deemed 'reliable'. With respect to the machine-approved sample, the human-rated sample is in all cases between roughly a factor of $\sim$1 to 2 smaller in each bin relative to the features examined in

Figure 4.14. In general the shapes of the machine-approved and human-approved distributions match fairly closely. Figure 4.15 shows examples of lightcurves for which the machine- and human-based classifiers differed in their rating (we focus on very short and very long fitted periods in Figure 4.15, but many examples exist for intermediate periods as well).

4.5.4 Asteroids with multiple fitted periods

A total of 654 unique asteroids have more than one PTF lightcurve whose fitted period was labeled as reliable by the vetting process described in the previous sections. These 654 asteroids collectively have 1,413 fits (so that the average multiplicity is $\sim$2.2 fits per asteroid) and include objects either observed in multiple oppositions and/or in both filters during one or more oppositions. Figure 4.16 plots the distribution of the relative error in the fitted frequencies of all such multiply-fit asteroids, this error being defined as the *range* of the asteroid's fitted frequencies divided by the *geometric mean* of its fitted frequencies. Just as in Figure 4.5 (when we compared to literature-referenced frequencies), we see a prominent mode in the histogram peaking at $\sim$0.1% relative error, with some excess for errors greater than $\sim$3% error. There are 63 asteroids in particular with relative errors greater than 3%, of these only four asteroids have more than two fits. If we assume that, in the remaining 59 pairs of disagreeing periods, one of the periods is correct, then the contamination fraction of lightcurves based on the sample of multiply-fit asteroids is $\sim 30/1413 = 4\%$.

4.6 Preliminary lightcurve-based demographics

In this section we perform a *preliminary* analysis and interpretation of the demographic trends evident in this work's fitted lightcurve parameters. Forthcoming works and papers will more closely examine the population distributions of both rotation and phase-function parameters.

Throughout this section we repeatedly examine variation of lightcurve-derived parameters as functions of *color index* and *infrared-derived diameters*. In the appendix

we describe the aggregation and characteristics of these two custom data sets (compiled from external sources). The color index quantifies an asteroid's probability of membership in the C-type ($p = 0$) versus S-type ($p = 1$) color-based clusters. Objects which in fact belong to neither C nor S groups (e.g., V types, D types) will have color indices near $p = 0.5$ provided they are in fact separated from both the C-type and S-type clusters in the 2D color spaces considered (see appendix).

There are many interesting demographic questions addressable with these lightcurve data which—in the interest of space—we do not treat in this work. For example, one could examine relationships between lightcurve parameters and orbital elements and/or family membership, proximity to resonances, and so on. We are making all of these lightcurve data available electronically (Tables 4 and 5, cf. Section 4.9.3) so that the community may use these data to help explore such science questions.

4.6.1 Disclaimer regarding de-biasing

The preliminary demographic analyses that follow do *not* take into account fully de-biased distributions of, e.g., spin rates, amplitudes, or phase-function parameters. The true-positive and false-positive rates given in the bottom row of plots in Figure 4.7 and 4.8 (also, the blue and violet lines in Figure 4.14), constitute some of the necessary ingredients for producing a fully de-biased data set, however in this work we do not attempt to compute the de-biased distributions.

4.6.2 Rotation rates and amplitudes

In Figure 4.17 we reproduce several of the plots appearing in [Pravec et al. (2002)] and references therein, using this work's much larger data set (characterized by at least an order of magnitude larger sample of small objects). Both spin rate and amplitude are examined for the 4,040 objects having diameter data from infrared surveys. Unlike [Pravec et al. (2002)], we are not able to individually plot each lightcurve's data (the ∼4,000 points would make the plot difficult to render, as well as difficult to read); hence we plot these (and other relationships later in this section) using two-

dimensional histograms where the intensity of each pixel corresponds to the number of objects in that bin (darker means more, with linear scaling). Additionally, 2D histograms for which the diameter is plotted on the horizontal axis have their pixel values column-normalized, i.e., all pixels in each column of the histogram sum to the same value. This facilitates the visual interpretation of period and amplitude variation with diameter, as the left-hand side (small-diameter end) of the plots would otherwise saturate the plot.

Following [Pravec et al. (2002)], we include the geometric mean rotation frequency as computed from a running bin centered on each object. The half-width of the bin centered on each object is either 250 (data points) or the object's distance from the top or bottom of the sorted diameter list, whichever is smallest. This ensures the geometric mean is not contaminated at the edges of the plot by the interior values, though it also means more noise exists in these edge statistics. The geometric mean is the more intuitive statistic for the rotation period as compared to the *arithmetic* mean, since the rotation periods tend to span several orders of magnitude. In addition to the geometric mean, we plot the 16th and 84th percentile values from each running bin.

The basic observed trend regarding rotation rate is that smaller-diameter asteroids rotate faster on average. A slight increase in the rotation rate also appears for objects larger than ∼80 km. Binning the data into a coarser set of three diameter bins and normalizing each object's spin rate by the local geometric-mean rate, we see a progression from a near-Maxwellian distribution to a progressively non-Maxwellian distribution for smaller objects. The rotation rates of a collisionally-equilibrated population of rotating particles is known to approach that of a Maxwellian distribution (e.g., [Salo (1987)]), which for a population of N objects as a function of rotation frequency f is:

$$n(N, f, f_{\rm peak}) = \frac{4Nf^2}{\sqrt{\pi} f_{\rm peak}^3} \exp\left(-\frac{f^2}{f_{\rm peak}^2}\right), \tag{4.32}$$

where $n(N, f, f_{\rm peak})df$ is the number of objects in the interval $(f, f + df)$ and $f_{\rm peak}$ is

the peak frequency (i.e. the frequency corresponding to the distribution's maximum).

One way of testing how well a Maxwellian actually fits the data is the two-sided Kolmogorov-Smirnov (KS) test [Massey (1951))]. This test compares an empirical distribution to a reference distribution (e.g., Gaussian, Maxwellian, or another empirical sample) via a bootstrap method. In particular it computes a statistic quantifying the extent to which the cumulative distribution function differs in the two distributions being compared. In our case, we use Equation (4.32) to simulate a large sample (10^5) randomly drawn from an ideal Maxwellian distribution and compare this simulated sample against the 99-asteroid sample (of $D > 40$ km) frequencies. Interestingly, this test indicates our 99 large-asteroid normalized frequencies differ from a Maxwellian at nearly the 10σ confidence level, though this could be due in part to the lack of a proper de-biasing of the distribution (cf. Section 4.6.1)

All of these trends—including the qualitative resemblance of a Maxwellian but its formal disagreement—were noted previously by [Pravec et al. (2002)]. At the time their $D < 10$ km size bin contained data on only 231 objects, as opposed to our sample of 2,844 asteroids with $D < 10$ km. Conversely, our $D > 40$ km bin contains only 99 objects as compared to the ∼400 large asteroids they took into consideration in comparing to a Maxwellian.

[Steinberg & Sari (2015)] recently described how collisional evolution of large asteroids should actually lead to a Lévy distribution, which has a significantly longer tail than a Maxwellian distribution having the same peak. They compared their theory to spin rates of $D \geq 10$ km asteroids from the LCDB and found in general that the Lévy distribution fails to fit the spin distribution of large asteroids, suggesting that there may be a significant primordial component to the spin distribution. Potential primordial contributions to the angular momentum of asteroids were explored by [Harris & Burns (1979)] and later authors; we will return to this topic in a later section.

Our amplitude distribution contains an obvious observational bias (cf. Section 4.6.1) in that amplitudes less than ∼0.1–0.2 mag are generally ill-fit by our modeling procedure (cf. Figure 4.7) and thus significantly underrepresented in our sample of

reliable lightcurves considered here. Nonetheless, we see a clear trend of smaller asteroids exhibiting larger rotational amplitudes, consistent with the idea that larger bodies have sufficient surface gravity to redistribute any loose mass to a more spherical shape.

As we have done for the normalized frequency distribution, we plot diameter-binned normalized amplitudes against a Maxwellian distribution, this time merely to guide the eye as opposed to validating any hypothetical physical interpretation. The fact that the normalized amplitude distributions do not deviate too drastically from the Maxwellian shape at smaller diameters indicates that the spread in the amplitude distribution is proportional to its mean value, a basic property of the Maxwellian distribution, hence the good agreement. [Carbognani (2010)] provides a recent analysis of asteroid rotation amplitudes, and highlighted a similar increase in both the amplitude's mean and spread with decreasing diameter.

Panel B of Figure 4.17 shows the distribution in period-vs.-amplitude space, in which we can plot *all* 9,033 lightcurves, including those lacking a diameter estimate. Contours representing the maximal spin rate of a body held-together solely by self-gravity of certain uniform densities are overplotted. Our data as a whole do not appear to populate the region beyond the $\sim$2 g/cm^3 contour. Later in this section we will re-examine this behavior separately for the two major taxonomic classes.

4.6.3 Phase-functions and bond albedos

We consider any of the 54,296 fitted PTF lightcurves to have a reliably-fit phase function if *both* of the following conditions are satisfied:

1. The lightcurve is one of the 9,033 having a reliable period fit, *or* its fitted amplitude (for the G_{12} model) is less than 0.1 mag (the latter is true for 1,939 lightcurves, only 39 of which have reliable periods)

2. The lightcurve is fit using data from at least five phase-angle bins of width $\Delta\alpha = 3$ deg. These five bins need not be contiguous, and they need not include

phase angles in the region where opposition surges are typically measured (i.e., $\alpha \lesssim 10$ deg)

The above two criteria are met by 3,902 out of the 54,296 PTF lightcurves. Of these, 1,648 have an infrared-based diameter available, 651 have a color index available, and 361 have both a diameter and color index.

Figure 4.18 details the distributions of the fitted phase parameters G_{12}, G, β and C against the color index, bond albedo, and in 1D histograms with color-based taxonomic subsets. Though the phase parameters are all correlated with color index and with bond albedo, none of the 1D phase-parameter distributions (right column of plots) exhibit bimodality alone, whereas the bond albedo (bottom right plot) does show significant bimodality. The red and blue histograms consist of all asteroids having color metric either less than 0.25 (C types) or greater than 0.75 (S types). The G and (β, C) phase parameters are only plotted for those lightcurves which also have a G_{12} solution. Not every lightcurve produced a solution for all three of the phase-function models, hence the sample sizes for the G and (β, C) models include a slightly reduced number of data points.

We reiterate our statement from Section 4.3.2.1 that the bond albedo $A_{\rm bond}$ is a more fundamental (i.e., intensive rather than extensive) property than is the geometric albedo p_V, hence our focus on $A_{\rm bond}$ here. The bond albedo is computed using Equation (4.8) together with Equation (4.15), and makes use of our PTF-derived absolute magnitudes—H from the G_{12} fit in particular—as well as the phase integral q of Equation (4.8), also computed directly from the G_{12} fit for ϕ. In particular,

$$q(G_{12}) = \begin{cases} 0.2707 - 0.236 G_{12} & \text{if } G_{12} < 0.2; \\ 0.2344 - 0.054 G_{12} & \text{otherwise.} \end{cases} \tag{4.33}$$

4.6.3.1 Taxonomy from lightcurve data

We use the distribution of bond albedo versus G_{12} to define another taxonomic metric analogous to the color index. In particular, we apply the same clustering analysis to this distribution as we did for the seven 2D color distributions in the appendix.

This procedure assigns to every object in the $A_{\rm bond}$-vs.-G_{12} diagram a probability of membership in each of two clusters (color coded blue and orange in Figure 4.19). The cluster centers are fit by the algorithm, and the output class probability of a given data point relates to its distance from these cluster centers. Probabilities near 0 represent likely C-type class membership, while probabilities near 1 represent likely S-type membership. We refer to this new metric as the *photometric index*; it complements the color index as another proxy for taxonomy. There are 361 asteroids with both a photometric index and color index available (Figure 4.19 right plot); the two indices are clearly correlated ($\rho_{\rm Spearman} = 0.73$, $>10\sigma$ significance). Note that asteroids only have a defined photometric index if they have an infrared-derived diameter available, so that $A_{\rm bond}$ is defined.

4.6.3.2 Wavelength dependence

Observational evidence for the reddening of asteroid colors with increasing phase angle is discussed by [Sanchez et al. (2012)] and references therein. Color variation with phase angle can be equivalently stated as variation of the phase function with wavelength. Asteroids which have PTF lightcurves in both of the survey's filters (R and g band) allow us to investigate this phenomenon. We note however that [Sanchez et al. (2012)] describe phase reddening as being more pronounced at longer wavelengths (>0.9 μm) and larger phase angles ($\alpha > 30$ deg), such that *a priori* we should not expect a very pronounced effect (if any) in the visible band PTF data.

Similar to the complication associated with comparing spin amplitudes from multiple oppositions (Section 4.3.1.1), an asteroid's mean color can potentially change if the spin axis varies with respect to our line-of-sight from year to year. Hence, we choose not to compare R-band and g-band phase-function fits from different oppositions. Aside from this constraint, we adopt the same two reliability selection criteria stated in Section 4.6.3, with a slight modification of requirement #2: here we allow *four* or more phase-angle bins of width $\Delta\alpha = 3$ deg, as opposed to the previous sections' five-bin requirement, because of the small sample size.

There are 92 asteroids with both R-band and g-band phase-function fits acquired

during the same opposition that meet the above criteria. For each asteroid we difference the R-band G_{12} value from the g-band G_{12} value. The mean of this difference is $-0.004^{+0.19}_{-0.14}$, indicating (for the whole sample) no significant non-zero difference between the two bands' G_{12} values. Likewise, for β, we compute a difference of $0.002^{+0.008}_{-0.003}$, also consistent with zero difference between the bands.

Since these fits provide absolute magnitudes in each band (i.e., H_g and H_R) we compute the color $H_g - H_R$ for the 92-asteroid sample. Figure 4.20 shows that the distribution of this color is bimodal, suggesting it is a viable proxy for taxonomy. This is further supported by the strong correlation between $H_g - H_R$ and the R-band G_{12} value. No correlation is seen however between $H_g - H_R$ and the difference between the two bands' G_{12} value or β values.

4.6.4 Spins and amplitudes vs. taxonomy

The union of the color-index data (see appendix) and photometric-index data (Section 4.6.3.1) provides significantly better taxonomic coverage of the PTF lightcurves (Figure 4.21). With this composite taxonomic information in hand, we can repeat the spin-amplitude analyses of Section 4.6.2 (Figure 4.17), this time considering the C-type and S-type groups separately. We define objects with one or both of the indices less than 0.25 as C type and greater than 0.75 as S type. We detail the resulting 1,795-object taxonomically-classified sample in Figure 4.22. There were 20 asteroids with conflicting color-based and photometric-based classifications that are not included in this 1,795-object sample.

The one-dimensional histogram in Figure 4.22 indicates that S-type asteroids dominate the smallest objects with data in PTF while C type dominate the largest. This reflects the fact that the survey's upper and lower sensitivity limits are defined in terms of absolute magnitude H (affected by albedo) rather than physical diameter, i.e., S-type asteroids larger than $\sim$50 km will tend to saturate the PTF detector, while C-type asteroids fainter than $\sim$5 km will usually fall below the detection limit. Adding to this effect is the fact that S-types mostly occupy the inner main-belt,

where they are brighter by virtue of smaller heliocentric and geocentric distances, as compared to the usually more distant C types. While the two classes have similar representation in the sample (882 S types versus 913 C types), their true population ratio also affects the relative numbers.

The right-hand side plots in Figure 4.22 show rotation rate and amplitude versus diameter separately for the two taxonomic groups. Rather than plot a two-dimensional histogram as was done in Figure 4.17, for readability we here just plot the geometric mean and percentiles, computed by exactly the same running-bin method described in Section 4.6.2. The most prominent trend is that among $5 \lesssim (D/\mathrm{km}) \lesssim 20$ asteroids, C types appear to rotate slower than S-types *and* have larger amplitudes than S types. Assuming both asteroid groups share the same mean angular momentum, the discrepancy could reflect the C types' ability to more efficiently redistribute material away from their spin axis, thereby increasing their moment of inertia (amplitude) while decreasing their angular rotation rate (i.e., a simple manifestation of conservation of angular momentum).

The above-stated *assumption* of a common mean angular momentum between C and S types is a merely a simple case and is neither unique nor rigorously motivated. More careful consideration of, e.g., plausible ranges of internal tensile strengths of the two types could easily lead to more diverse scenarios wherein the two groups actually have different angular momenta and the observed spin-amplitude trends. As noted earlier (Section 4.6.2), large asteroids in general appear to have retained a significant primordial component in their spin distribution [Steinberg & Sari (2015)]; it is therefore important that differences in the origin of C types and S types (accretionary, temporal and/or spatial) be taken into account along with differences in collisional evolution and differing contributions from radiative forces like YORP. Simulations of the main belt's origin, such as the Grand Tack family of models [Walsh et al. (2011)], should ultimately be modified to track particle spin evolution as well as orbits.

We also reproduce the period-vs.-amplitude plot first shown in Figure 4.17, this time plotting separately the two taxonomic groups. The S types show a clearer cutoff at the 2 g/cm^3 contour line, suggesting they may in general be of greater bulk density

than the C types, which show a softer boundary in this period-vs.-amplitude space, the precise location of which appears to be somewhere between 1 and 2 g/cm^3. Note that comparison to these density contours is only valid if the asteroids in consideration are held together mostly by self-gravity and approximated as fluids (as opposed to having significant internal cohesive or frictional resistance). These results are in general agreement with existing asteroid density estimates ([Carry (2012)] and references therein). Results from an independent analysis of a smaller, more densely-sampled set of PTF asteroid lightcurves (Chang et al. in review; a study that follows closely the approach of [Chang et al. (2014a)]) agree with the C type vs. S type rotation rate discrepancy discussed here.

4.7 Comparison to MPC-generated magnitudes

Absolute magnitudes available through the Minor Planet Center (MPC) and JPL Solar System Dynamics[15] websites are fit using all available survey/observer-contributed photometry. These H magnitudes are used in various online ephemeris tools to compute predicted V magnitudes to accompany astrometric predictions. Their model assumes no rotational modulation, uses the Lumme-Bowell G-model (Section 4.3.2.2), and—with the exception of $\sim$100 large objects (nearly all with $D > 30$ km)—assumes a constant $G = 0.15$ for all asteroids. Our results (Figure 4.18 second row of plots) show that the $G = 0.15$ approximation does indeed agree well with the peak of the distribution of fitted G values. The PTF-fitted G values obviously however show some spread and variation with taxonomy. In this section we explore the resulting differences in the absolute magnitudes H and in predicted magnitudes.

4.7.1 Filter transformations

In order to compare the MPC-listed ($H_{\rm MPC}$) magnitudes, which are in V band, with PTF's absolute magnitudes ($H_{\rm PTF}$, corresponding to the G-model fit) which

[15]http://ssd.jpl.nasa.govhttp://ssd.jpl.nasa.gov

are in either R and g bands, we must first compute an approximate transformation from V-band to each PTF band. While some transformations are given by [Ofek et al. (2012a)], we here prefer to empirically estimate these using actual asteroid photometry from both PTF and the MPC, rather than generating them from the more general transformations of [Ofek et al. (2012a)].

Figure 4.23 plots $H_{\rm PTF} - H_{\rm MPC}$ for asteroids whose PTF-derived $G_{\rm PTF}$ is in the range $0.1 < G_{\rm PTF} < 0.2$. By restricting the comparison to objects with fitted $G_{\rm PTF}$ values close to 0.15, we in principle select $H_{\rm MPC}$ magnitudes for which the MPC's $G_{\rm MPC} = 0.15$ assumption is actually valid (none of the asteroids in Figure 4.23 have MPC-listed G values other than the default 0.15). Furthermore, we only consider (in Figure 4.23) asteroids with PTF data in at least three phase angle bins of $\Delta\alpha = 3$ deg *and* either a reliable period or fitted amplitude less than 0.1 mag.

Comparing the $H_{\rm MPC}$ and $H_{\rm PTF}$ magnitudes for this specific subset of asteroids, we obtain approximate transformations $R = V + (0.00 \pm 0.10)$ and $g = V + (0.55 \pm 0.16)$. The 1σ uncertainties of 0.10 and 0.16 mag plausibly include a combination of the photometric calibration uncertainties of both the MPC data (coming from a variety of surveys/observers), variation in H magnitude of a given asteroid between different oppositions (the MPC fits combine data possibly acquired at different viewing geometries), as well as the range of $G_{\rm PTF}$ used in selecting the asteroids in this sample. Consideration of a range of $G_{\rm PTF}$ values is equivalent to considering a range of asteroid colors (cf. the color-vs.-G correlation seen in Figure 4.18). Hence the uncertainties in these transformations also encompass the variation which might otherwise be formally fit in a color term for the transformations. Such a color term for R to V would almost certainly be less significant than that of g to V, as the former transformation is already zero within uncertainties. The larger uncertainty in the g to V transformation is likely attributable to both the smaller sample size and the fact that the V bandcenter is further displaced from g than from R, such that color variation has a more pronounced effect.

Given the above-computed transformations, and the fact that 89% of our fitted lightcurves are in R band, we proceed using only R-band lightcurve fits, which we

compare directly against MPC magnitudes (or formally, after applying the transformation of zero). A detail of the color dependence of the R to V transformation appears in the right plot of Figure 4.23; the mean transformation differs slightly between S and C types but not at a level comparable to the uncertainty in either.

4.7.2 Absolute magnitudes

In Figure 4.24 we show the relative error in the MPC absolute magnitudes as compared to the PTF magnitudes, for all 1,630 lightcurves with sufficient phase angle coverage in PTF (with the five-bin phase-angle criterion). These errors should reflect not only any discrepancy due the different phase function models (PTF's G_{12} versus MPC's G), but also variation in absolute photometric calibrations (within the MPC data internally and/or between the MPC and PTF data sets). The 0.1-mag uncertainty in the R to V band transformation has a prominent contribution to the errors shown here (the mean and 84th percentile of the errors expected from the 0.1-mag transformation uncertainty alone are shown as yellow dashed lines, and assume $p_V = 0.07$). The green line (computed mean) and upper red line (84th percentile) indicate the errors are $\sim$1% greater than those expected from the transformation uncertainty alone, though this increases slightly for the largest ($D > 30$ km) objects. Note that many of these largest asteroids are more frequently observed by programs other than the major sky-surveys; these smaller facilities tend to use smaller aperture telescopes and different absolute calibration standards, which would contribute to the error.

4.7.3 Predicted apparent magnitudes

Instead of comparing just the fitted H magnitudes, for every lightcurve with a reliable PTF-fitted phase function we also compare the root-mean-square residual of all PTF data in that lightcurve with respect to both our G_{12}-fit-predicted R magnitude and the MPC ($G = 0.15$) predicted V magnitude. Our fit includes more fitted parameters and obviously should result in smaller residuals; Figure 4.25 shows that we see a

factor ∼3 smaller residuals in particular using the PTF fit. Note that if the 0.1-mag R-to-V transformation uncertainty were the only significant contributor to the MPC residuals then their peak would instead be at ∼0.07 mag rather than ∼0.25 mag. Ignored rotational modulation and inaccurate phase functions move the MPC residuals distribution to higher RMS values.

The lower RMS residuals afforded by the PTF lightcurve model permit a more sensitive search for low-level transient activity (e.g., collisional events, cometary brightening) in these asteroids. For example, [Cikota et al. (2014)] perform a search for active main-belt asteroids using photometric residuals of all MPC data taken with respect to the MPC-predicted apparent V magnitudes. We currently are pursuing a similar analysis using these PTF lightcurves, as a follow-up to the morphology-based search already completed with PTF [Waszczak et al. (2013b)]. A hybrid approach, wherein morphological measurements are made on stacked images of asteroids which have reliable lightcurve fits, could further reveal this kind of subtle activity.

4.8 Summary

From five years of PTF survey data we have extracted over 4 million serendipitous detections of asteroids with known orbits. We fit a photometric model to ∼54,000 lightcurves, each consisting of at least 20 observations acquired within a given opposition in a single filter. We adopt a second order (four-term) Fourier series for the rotation component and fit three distinct phase-function models. We assess the reliability of our retrieved rotation periods by subjecting them to both an automated classifier and manual review. Both vetting processes are trained on a sample of ∼800 asteroids with previously measured spin periods that also occur in our sample. We consider the intersection of the two screened samples for subsequent analysis.

Preliminary analysis (on distributions that are not de-biased) of the rotation period versus diameter confirms the previous finding that asteroids smaller than ∼ 40 km do not conform to a Maxwellian distribution in their normalized spin frequencies. Phase-function parameters are shown to correlate strongly with the bond albedo.

None of the phase function parameters display bimodality in their measured distributions however. Together with the bond albedo, we use the phase function data to define a new taxonomic metric based solely on single-band lightcurve properties together with infrared-derived diameters (G_{12} and $A_{\rm bond}$). This metric complements the color-based index established previously by many visible-color and spectroscopic surveys. Combining these color- and photometry-based taxonomic indices allows us to separately examine the spin and amplitude distributions of the C-type and S-type asteroids with the largest possible sample sizes. Doing so reveals that, among small objects ($5\ {\rm km} < D < 20\ {\rm km}$) the C types show larger amplitudes and slower spin rates. If the two populations shared a common angular momentum distribution, this could be interpreted as the two compositional types' differing tendencies to redistribute mass away from their spin axes. Comparison of the spin-amplitude distribution with contours of maximal spin rates for cohesionless bodies suggests that almost all asteroids are less dense than $\sim$2 g/cm^3, with C types displaying a potentially less dense upper limit of between 1–2 g/cm^3.

Finally, our fitted absolute magnitudes differ from those generated by the Minor Planet Center's automated fitting procedures, though the precise discrepancy is difficult to ascertain given uncertainty in the transformation between PTF R-band and the MPC's V-band. The utility in using our model to *predict* asteroid apparent magnitudes is seen in the three-fold reduction in RMS scatter about our model relative to the fiducial $G = 0.15$ model that neglects rotation. This reduced scatter is an essential prerequisite for sensitive searches for cometary, collisional, and other transient activity in what would otherwise be regarded as quiescent asteroids—potentially even bright objects.

Table 4.3: Asteroid colorimetry data sets used in computing this work's C/S color metric. These data sets are visualized in Figure 4.26.

survey name	references	data description	# asteroids
UBV colors	[Bowell et al. (1978)] [Tedesco (1995)]	U, B, and V broadband photometry acquired mostly at Lowell Observatory in the 1970s with photomultiplier tubes.	902
Eight-Color Asteroid Survey (ECAS)	[Zellner et al. (1985)] [Zellner et al. (2009)]	Photometry in eight custom filters measured with photomultipliers at Catalina and Steward Observatories. We compute and use the principal component color index PC#1 $= 0.771(b-v) - 0.637(v-w)$. Excludes objects with PC#1 error >0.3 mag.	480
24-Color Asteroid Survey	[Chapman & Gaffey (1979)] [Chapman et al. (1993)]	Photometry in 24 interference filters measured with photomultipliers at Mauna Kea. We compute and use the mean spectral reflectance slope and first principal component.	262
Small Main-belt Asteroid Spectroscopic Survey (SMASS)	[Xu et al. (1995)] [Xu et al. (1996)]	CCD spectroscopy (0.4–1.0μm, $R \approx 100$) conducted mostly at Kitt Peak. We compute and use the mean spectral reflectance slope and first principal component.	305
Small Main-belt Asteroid Spectroscopic Survey II (SMASS-2)	[Bus & Binzel (2002)] [Bus & Binzel (2003)]	CCD spectroscopy (0.4–1.0μm, $R \approx 100$) conducted at Kitt Peak. We compute and use the mean spectral reflectance slope and first principal component.	1,313
Small Solar System Objects Spectroscopic Survey (S3OS2)	[Lazzaro et al. (2004)]	CCD spectroscopy (0.5–9.0μm, $R \approx 500$) conducted at ESO (La Silla). We compute and use the mean spectral reflectance slope and first principal component.	730
Sloan Digital Sky Survey (SDSS) $griz$ colors	[Ivezić et al. (2002)] [Parker et al. (2008)] [Ivezić et al. (2010)]	g,r,i, and z broadband CCD photometry acquired by SDSS from 1998–2009. Includes data in the Moving Object Catalog v4, supplemented with post-2007 detections from SDSS DR10. We use the first principal component a^* defined in the references. Excludes objects with a^* error >0.05 mag or $(i-z)$ error >0.1 mag.	30,518

4.9 Appendix

4.9.1 Multi-survey visible-band color index

The purpose of this appendix section is to introduce a one-dimensional color metric, based upon data from seven different colorimetric asteroid surveys, which quantifies an asteroid's first-order visible-band color-based taxonomy as a number between 0 (C-type endmember) and 1 (S-type endmember). Our primary motivation for doing this is to enable a uniform comparison of PTF-lightcurve-derived parameters with color spanning from the brightest/largest objects ($H \approx 8$–9 mag, or $D \approx 125$–80 km diameters) down to PTF's detection limit for main-belt asteroids ($H \approx 16$ mag, or $D \approx 2$–4 km). Figure 4.26 panel A shows that the fraction of PTF lightcurves with color information increases by a factor of $\sim$3 among large asteroids when all seven surveys are considered, whereas for smaller objects the Sloan Digital sky-survey's (SDSS; [York et al. (2000)]; [Ivezić et al. (2002)]; [Parker et al. (2008)]) moving-object catalog provides essentially all of the color information.

The seven surveys we use are described in Table 4.3. All of these surveys contain at least two independent color measurements, and when plotting their data in these two dimensional spaces (or 2D subspaces defined by properly-chosen principal components or spectral slope parameters), the first-order C-type and S-type clusters are in all cases prominently seen (Figure 4.26 panel B). To each such 2D color distribution we apply a two-dimensional *fuzzy c-means* (FCM) clustering algorithm ([Bezdec (1981)]; [Chiu (1994)]). For each survey data set, FCM iteratively solves for a specified number of cluster centers (in our case, two) in N dimensions (in our case one dimension) by minimizing an objective function which adaptively weights each datum according to the robustness of its membership in a given cluster. The FCM output includes computed cluster centers and, for each datum, the probability that the datum belongs to each cluster (this being related to the datum's distance from each cluster center).

In the color-distribution plots of Figure 4.26 panel B (the plots with black backgrounds arranged diagonally), each pixel/bin is colorized according to the average cluster-membership probability of asteroids in that pixel. Blue indicates high prob-

ability of membership in cluster 1 while orange represents high probability of membership in cluster 2.

Our color index provides a more quantitative label than that offered by popular letter-based taxonomic systems (e.g., [Bus et al. (2002)] and refs. therein). Several such letter-based nomenclatures were in fact defined on the basis of one or more of these seven surveys, oftentimes using a method similar to the clustering technique we use here. We identify our blue cluster with C-type asteroids and our orange cluster with S-type asteroids, though we make this association purely for connection/compatibility with the literature. This is because our computed clusters have their own unique identity/definition, formally distinct from that given in any other work. Our clusters' definitions are nonetheless completely specified/reproducible by the FCM algorithm we used to compute them.

In reducing the taxonomic classification to a single number defined by the two most prominent groups (C and S types), we lose the ability to distinguish secondary classes like V types, D types, and so on. If such a sub-group is separated from both of the two main clusters, its members will be assigned membership scores of close to 0.5. For example, in the SDSS a^* vs. $i - z$ complex, the clearly-seen V-type 'tail' protruding down from the S-type cluster appears mostly green in color, reflecting its intermediate classification. Likewise for the less-clearly seen D types, which in the SDSS plot lie above the S types and to the right of the C types (again in a green-colored region). The orders of magnitude lower numbers of such secondary types make them mostly irrelevant for the purpose of this analysis.

We compute the numerical uncertainty (variance) of a given asteroid's cluster-membership score in a particular survey by performing many bootstrapped trials wherein we first randomly perturb all data points by random numbers drawn from Gaussian distributions whose width are the quoted 1σ measurement (i.e., photometric) uncertainties in each of the two dimensions, and then repeat the FCM analysis on the perturbed data. The variance in each object's reported cluster probability is then computed after a large number of bootstrap trials.

Some asteroids appear in only one of the seven surveys; for such objects the color

index is simply its cluster-membership score in that particular survey. For asteroids appearing in multiple surveys, we take the variance-weighted average of the multiple membership scores (and compute that composite score's variance by summing the component variances in inverse quadrature, as usual).

The many off-diagonal plots in Figure 4.26 panel B compare the cluster-membership scores of all asteroids appearing in all possible survey intersections. The number of asteroids in each survey (and in the intersection of each survey pair) appears above each plot ($N = \ldots$). The survey-pair distributions are 2D-histograms where higher densities of data points correspond to black pixels/bins and low density or lack of data points is white. Evidently all possible survey combinations contain at least some asteroids (several share hundreds), and in all cases the individual taxonomic indices (on the horizontal and vertical axes) correlate strongly, confirming the consistency of the cluster membership between surveys.

In Figure 4.27 we illustrate some useful applications of this color index by comparing it with various asteroid surface observations. One of these quantities (SDSS a^* color) was used in computing the color index, so its correlation with the clustering index is expected and thus confirmed.

In the leftmost plot of Figure 4.27, asteroid photometry from *GALEX*[16] (NUV band), compiled by Waszczak et al. (in prep), is normalized by the nominal $G = 0.15$ phase-model (Section 4.3.2.2) predicted brightness at the time of the *GALEX* observations, and the resulting NUV $- V$ color evidently correlates with the visible color index. This indicates that asteroid reflectance slopes in the visible persist into the UV.

Figure 4.27 also plots our color index against the $W1$-band geometric albedo derived from *WISE*[17] observations obtained during its fully cryogenic mission. We

[16]The *Galaxy Evolution Explorer* (*GALEX*) is a NASA Small Explorer-class space telescope which from 2003–2012 conducted an imaging survey in a far-UV band (FUV, 130–190 nm) and a near-UV band (NUV, 180–280 nm). [Martin et al. (2005))] discuss the extragalactic science program; Morissey et al. (2005, 2007) discuss the on-orbit performance, survey calibration and data products. The Waszczak et al (in prep) NUV data shown here are derived from data available at http://galex.stsci.eduhttp://galex.stsci.edu.

[17]The *Wide-field Infrared Survey Explorer* (*WISE*) is a NASA Medium Explorer-class space telescope which in 2010 conducted a cryogenic IR imaging survey in four bands: $W1, W2, W3$, and $W4$,

only include asteroids which were detected in both of the thermal bands ($W3$ and $W4$) and which therefore have a reliable diameter estimate. Use of this diameter in Equation (4.15) then permits estimation of the albedo, where the $W1$-band albedos uses the corresponding *WISE* photometry (H in Equation [4.15] being replaced with the appropriate $W1$-band absolute magnitude).

The rightmost plot in Figure 4.27 shows our color index's relationship to a near-infrared color from the ground-based 2MASS survey [Skrutskie et al. (2006)]. Serendipitous asteroid detections were extracted from 2MASS by Sykes ([Sykes et al. (2000)], [Sykes et al. (2010)]) and include fluxes in J band (1.25 μm), H band (1.65 μm—*not to be confused with the absolute visible magnitude H, used elsewhere in this work*), and K band (2.17 μm).

4.9.2 Compilation of IR-derived diameters

Similar to how we combined several surveys' colorimetric data in the previous section, here we compile thermal-infrared-derived diameter estimates from four surveys. Our aim is again to provide the largest possible sample for comparison with PTF-derived lightcurve data. Just as SDSS is the main contributor of colorimetry overall but suffers from incompleteness for large/bright asteroids, analogously *WISE* provides the vast majority of IR-based diameter measurements but levels off at ∼80% completeness at the bright end (Figure 4.28). We thus supplement *WISE* with diameter data from the *Infrared Astronomical Satellite* (*IRAS*; [Matson et al. (1986))], [Tedesco et al. (2002a)]), the *Mid-Course Space Experiment* (*MSX*; [Tedesco et al. (2002b)]), and AKARI [Usui et al. (2011)]. [Usui et al. (2014)] compares several of these different data sets in terms of coverage and accuracy. As we did when defining the color index, asteroids occurring in multiple IR surveys are assigned the variance-weighted average diameter.

Regarding the WISE data in particular, we again use only those diameters which resulted from a thermal fit constrained by fluxes in all four WISE bands during

centered at 3.4, 4.6, 12, and 22μm, respectively. [Wright et al. (2010)] details mission/performance; [Masiero et al. (2011)] and refs. therein present preliminary asteroid data.

the cryogenic mission. Furthermore, we use the latest (revised) diameter estimates published by [Masiero et al. (2014)], which adopted an improved thermal modeling technique first discussed by [Grav et al. (2012))].

4.9.3 Lightcurve data tables

The online version of this article includes two electronic tables containing the derived lightcurve parameters and the individual photometric observations in each lightcurve. Tables 4 and 5 describe the columns and formatting of these tables, which include data on all reliable-period lightcurves as well as those having amplitudes less than 0.1 mag and sampling in five or more 3-deg-wide phase-angle bins (which have reliable G_{12} fits). Using these tables one can produce plots of the PTF lightcurves we have analyzed in this work.

Table 4.4: Parameters describing PTF lightcurves with a reliable period or phase function. Byte-by-byte Description of file: `ptf_asteroid_lc_parameters.txt`

Bytes	Format	Units	Label	Explanations
1- 4	I4	—		Lightcurve ID number[1]
6- 11	I6	—		Asteroid number (IAU designation)
13- 14	I2	yr		Last two digits of opposition year
16	I1	—		Photometric band: 1 = Gunn-g, 2 = Mould-R
18- 20	I3	—		Number of observations in the lightcurve
22- 26	F5.2	mag		Median apparent magnitude
28- 37	F10.5	day	$t_{\rm min}$	Time (MJD) of first observation
39- 48	F10.5	day	$t_{\rm max}$	Time (MJD) of final observation
50- 54	F5.2	deg	$\alpha_{\rm min}$	Minimum-observed phase angle
56- 60	F5.2	deg	$\alpha_{\rm max}$	Maximum-observed phase angle
62- 63	I2	—		Number of sampled phase-angle bins of 3-deg width
65- 68	F4.2	—	p	Reliability score from machine classifier: 0=bad, 1=good
70	I1	—		Manually-assigned reliability flag: 0=bad, 1=good
72	I1	—		Period reliability flag: 0=bad, 1=good (product of two previous columns)
74- 79	F6.3	mag	H	Absolute magnitude from G_{12} fit
81- 85	F5.3	mag		Uncertainty in absolute magnitude from G_{12} fit
87- 91	F5.3	—	G_{12}	Phase-function parameter G_{12}
92- 98	F6.3	—		Uncertainty in G_{12}[2]
100-105	F6.3	—	G	Phase-function parameter G
107-113	F7.4	mag/deg	β	Phase-function parameter β
115-119	F6.3	mag	C	Phase-function parameter C
121-124	F4.2	mag		Amplitude from G_{12} fit (max − min)
126-134	F9.4	hr	P	Period from G_{12} fit
136-144	F9.4	hr		Period uncertainty from G_{12} fi
146-152	F7.4	mag	A_{11}	Fourier coefficient $A_{1,1}$ from G12 fit
154-160	F7.4	mag	A_{12}	Fourier coefficient $A_{1,2}$ from G12 fit
162-168	F7.4	mag	A_{21}	Fourier coefficient $A_{2,1}$ from G12 fit
170-176	F7.4	mag	A_{22}	Fourier coefficient $A_{2,2}$ from G12 fit
178-181	F4.2	—		Ratio of the two peak heights in folded rotation curve[3]
183-186	F4.2	—	$\chi^2_{\rm red}$	Reduced chi-squared of the fit
188-192	F5.3	mag		"Cosmic error" (see Section 4.4.1)
194-198	F5.3	mag		Root-mean-square residual of observations w.r.t the fit
200-206	F7.3	hr		Reference period (from http://sbn.psi.edu/pds/resource/lc)
208-213	F6.2	km	D	Diameter derived from thermal IR data[4]
215-218	F4.2	km		Uncertainty in diameter
220-224	F5.3	—	$A_{\rm bond}$	Bond albedo[5]
226-231	F6.4	—		Uncertainty in bond albedo
233-236	F4.2	—		Color-based taxonomic index: 0=C-type, 1=S-type
238-241	F4.2	—		Photometry-based taxonomic index: 0=C-type, 1=S-type

[1]ID number labels individual observations in Table 4.5.
[2]Set to −1 if larger than the interval tested in grid search
[3]Set to 0 if there is only one maximum in the folded lightcurve
[4]References for the IR diameters are given in the text (appendix)
[5]Bond albedo only computed for objects with reliable G_{12} and available diameter

Table 4.5: Parameters describing PTF lightcurves with a reliable period or phase function. Byte-by-byte Description of file: `ptf_asteroid_lc_observations.txt`

Bytes	Format	Units	Label	Explanations
1- 4	I4			Lightcurve ID number[1]
6- 15	F10.5	day	τ	Light-time-corrected observation epoch
17- 26	F10.7	AU	r	Heliocentric distance
28- 37	F10.7	AU	Δ	Geocentric distance
39- 43	F5.2	deg	α	Solar phase angle
45- 50	F6.3	mag	R or g	Apparent magnitude[2]
52- 56	F5.3	mag		Uncertainty in apparent magnitude
58- 62	F5.3	mag		Uncertainty in apparent magnitude with cosmic-error
64- 69	F6.3	mag		Magnitude corrected for distance and G_{12} phase function
71- 76	F6.3	mag		Magnitude corrected for distance and rotation (G_{12} fit)
78- 83	F6.3	mag		Residual with respect to the G_{12} fit
85- 89	F4.1			Rotational phase from 0 to 1 (G_{12} fit)

[1]ID number also corresponds to the line number in Table 4.4.
[2]Filter/band is specified in Table 4.4.

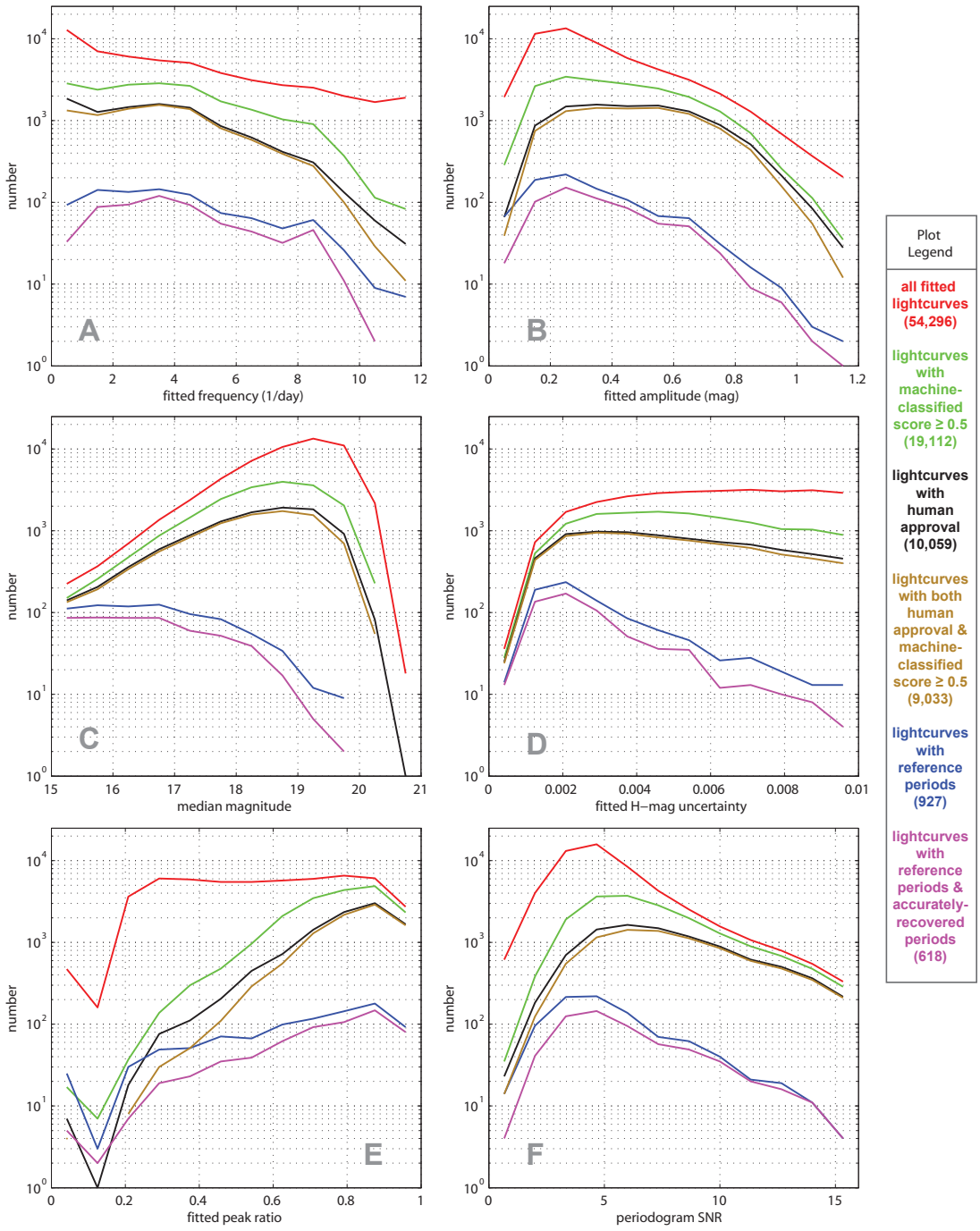


Figure 4.14: Distributions of PTF-fitted lightcurves (and various subsets thereof) in select features/parameters. These plots are histograms with the same binning as the top rows of Figures 7 and 8. For better readability we here use line-connected bin points (rather than the stair-plot format used in, e.g., Figure 4.5).

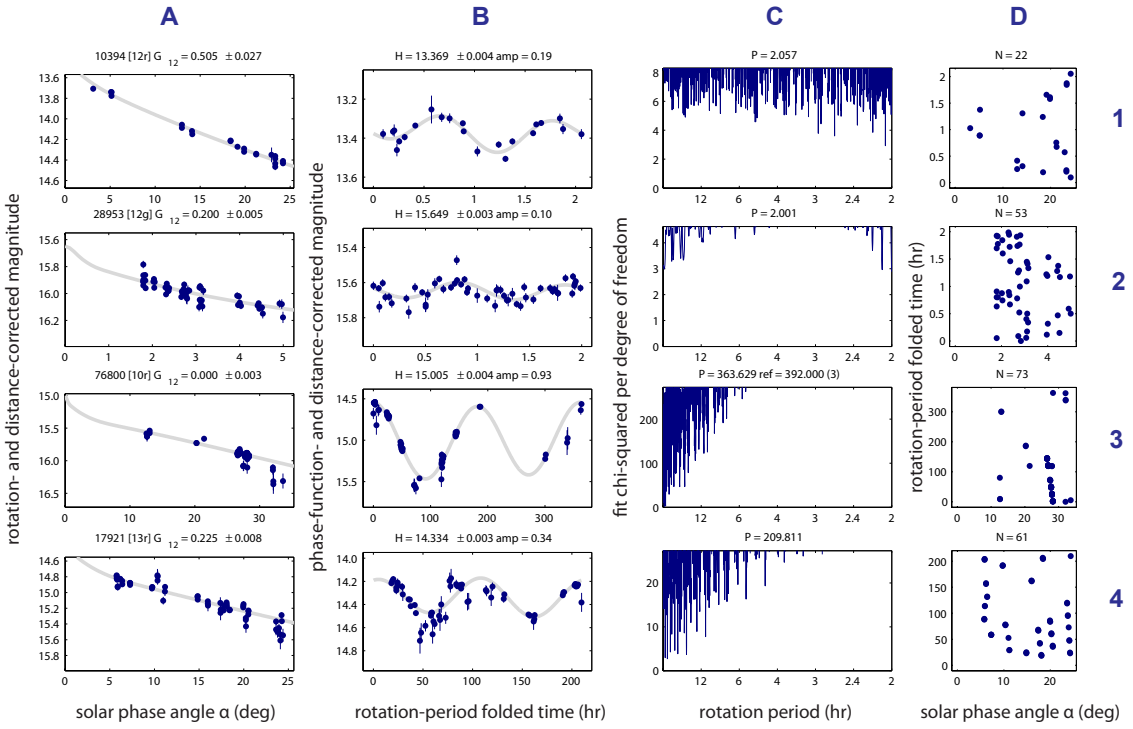


Figure 4.15: Example lightcurves for which the machine-based and human-based reliability scores differ. *Row 1*: Human approved, machine rejected ($p = 0.32$). *Row 2*: Human rejected, machine approved ($p = 0.66$). *Row 3*: Human approved, machine rejected. For this object, the fitted period differs from the known reference period of 392 hours by 7%, hence the machine rejects it by definition. *Row 4*: Human rejected, machine approved ($p = 0.70$).

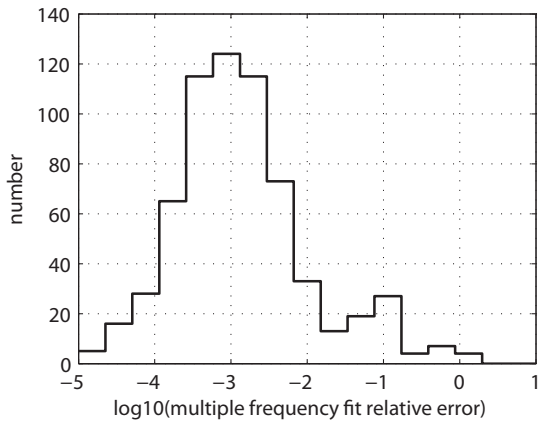


Figure 4.16: For the 654 unique asteroids having more than one reliable lightcurve fit (either multiple oppositions and/or both R and g band data) we plot the log of the relative frequency error, defined as the range of the asteroid's fitted periods divided by the geometric mean of its fitted periods. Comparison with Figure 4.5 suggests that we can deem all cases with error $\lesssim$3% as consistently recovered periods, and those with greater than 3% error as inconsistent fits.

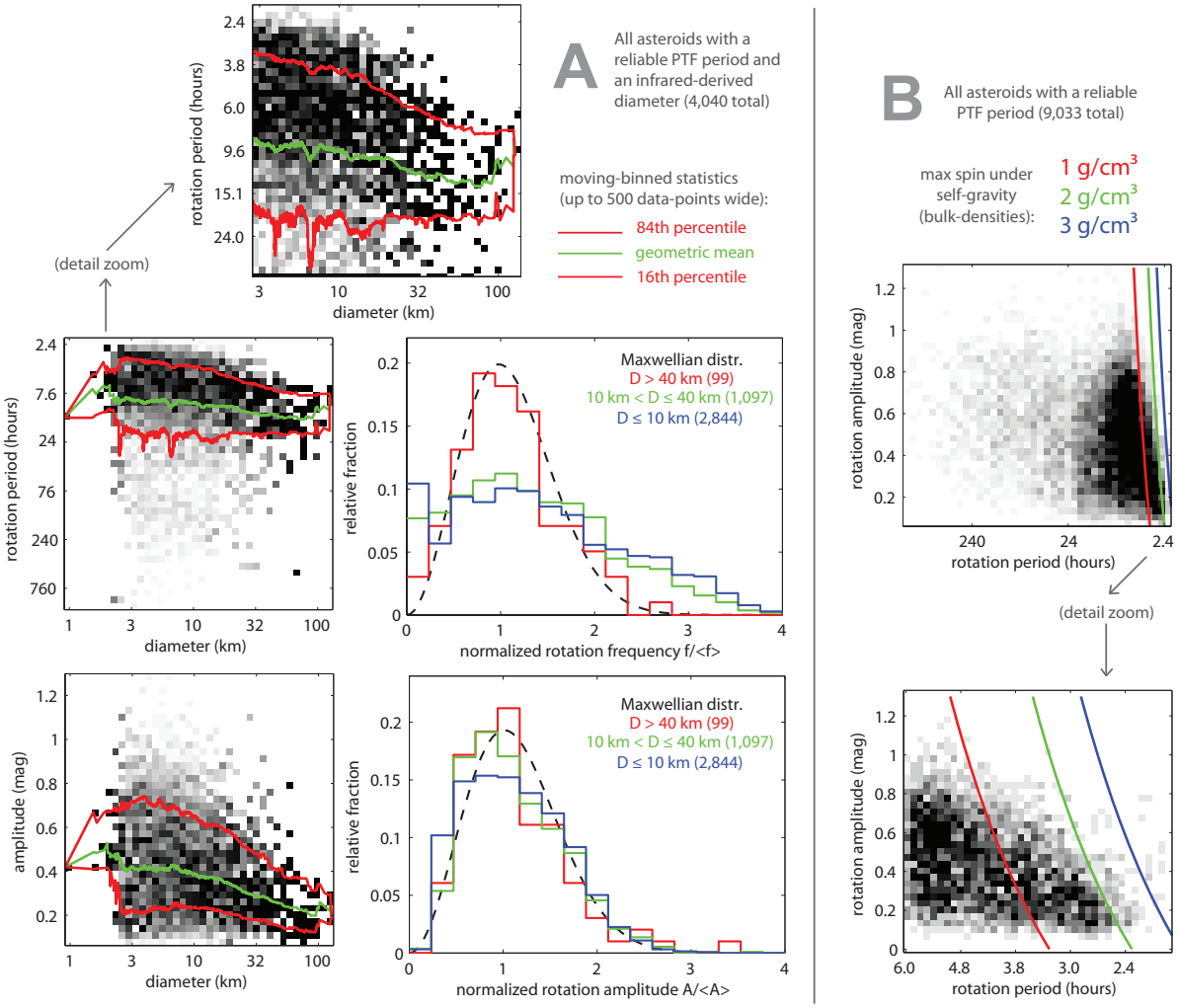


Figure 4.17: *Panel A*: Distribution of spin rate and amplitude as functions of infrared-derived diameters (see appendix for diameter data sources), including data for 4,040 of our lightcurves. The two-dimensional histograms (left side plots) are column-normalized (see text for details). *Panel B*: Comparison of the period versus amplitude distribution (regular 2D histogram, *not* column normalized) with max-spin-rate versus amplitude for a uniform density ellipsoid held together solely by self-gravity

.

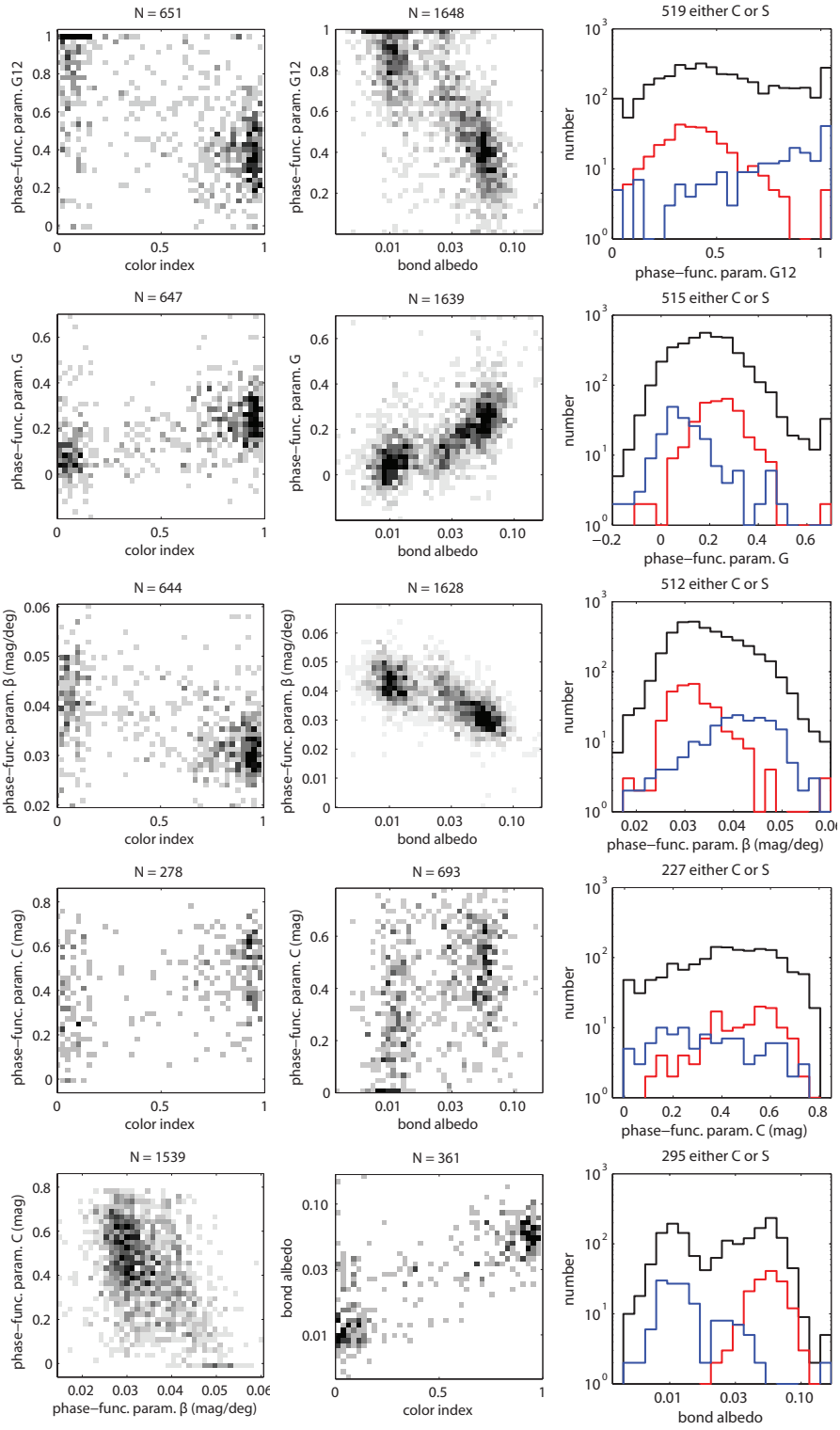


Figure 4.18: Various fitted phase-function parameters plotted against color index and bold albedo (two-dimensional histograms; the total number of lightcurves in each plot is stated above it as $N = \ldots$). In the right column of plots, one-dimensional distributions with the color-index classified objects plotted separately. In the right column of 1D histograms, C and S types are defined as objects with color indices less than 0.25 and greater than 0.75, respectively.

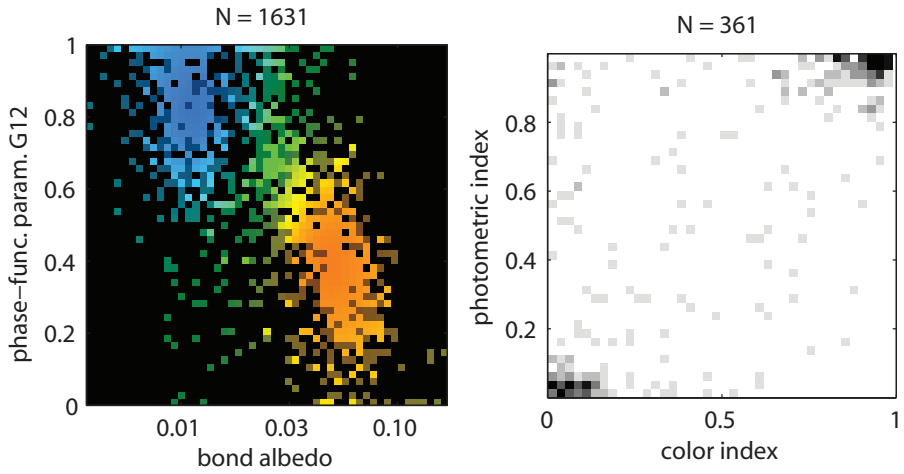


Figure 4.19: *Left*: We perform the same clustering analysis used in defining the color index (see appendix), this time on the G_{12} versus $A_{\rm bond}$ distribution, which contains 1,631 PTF lightcurves all of which have IR-derived diameters and reliable phase functions. The output of this clustering analysis is the *photometric index*, which analogous to the color index is a number between 0 (C type) and 1 (S type) quantifying to the class membership of each constituent asteroid data point. *Right*: Correlation between the color index and our photometric index, a comparison which can be made for 361 objects. Note that most data are in the lower left and upper right corners

.

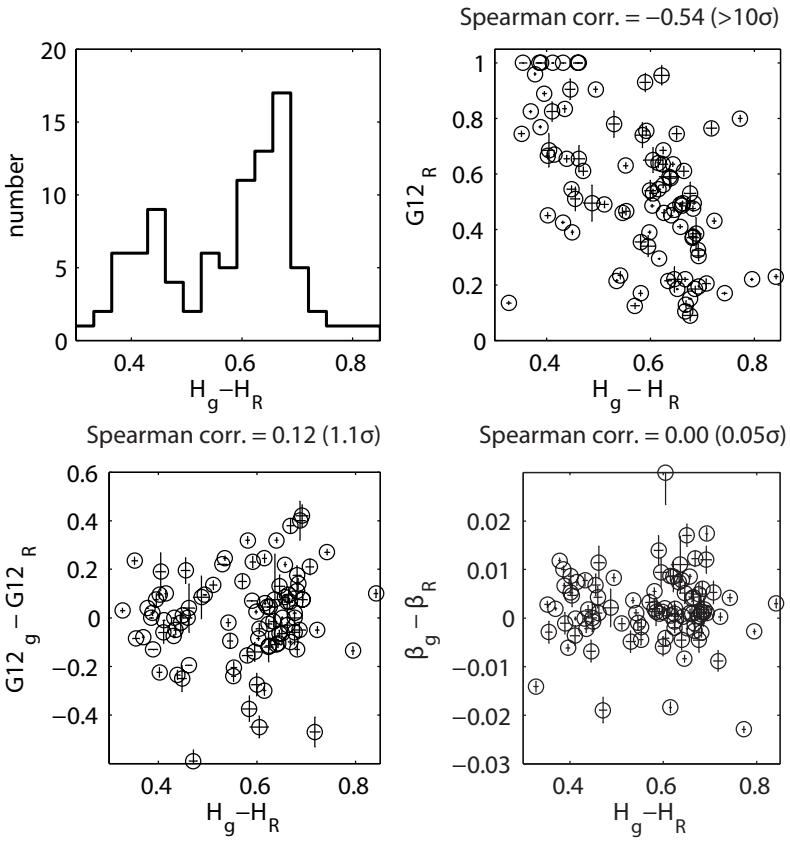


Figure 4.20: For the 92 asteroids with both R-band and g-band lightcurve fits from the same opposition, we use the resulting difference in the absolute magnitudes $H_g - H_R$ as a proxy for taxonomy. This color distribution is qualitatively bimodal (top left), and the correlation with G_{12} is very robust (top right). We detect no significant difference in the G_{12} and/or β parameters between the two bands, both in the sample as a whole, and as a function of the $H_g - H_R$ color.

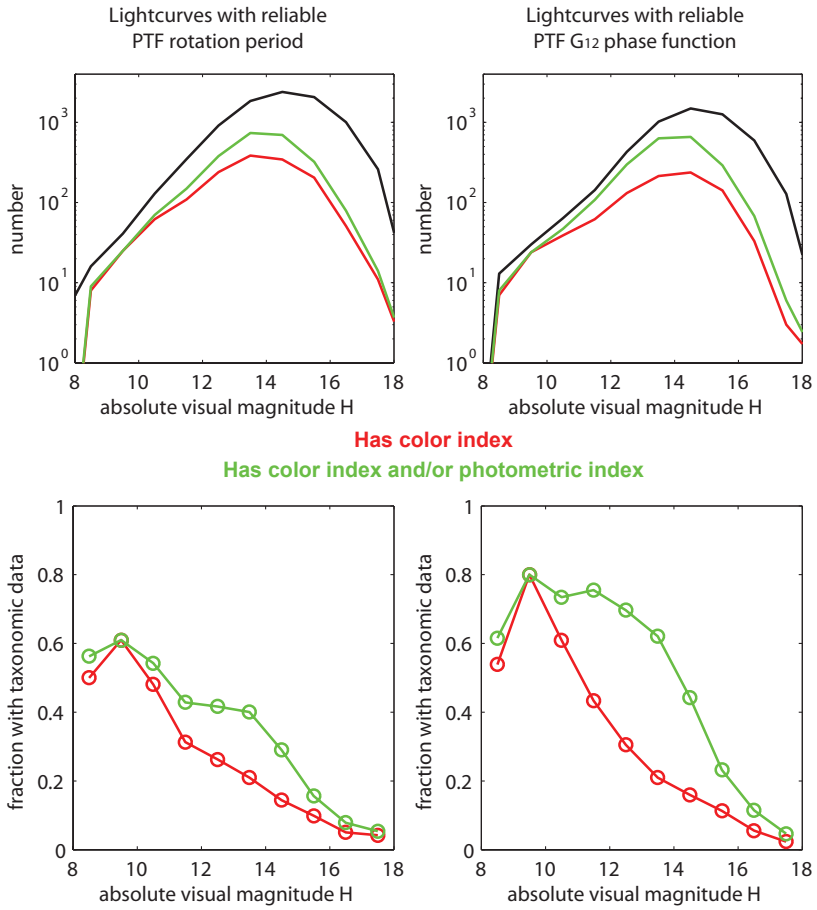


Figure 4.21: Added completeness from supplementing the color index with the photometric index among asteroids having PTF lightcurves. Both indices are a proxy for the taxonomic type. The left- and right-hand plots apply separately to the subset labeled by the black line above each column.

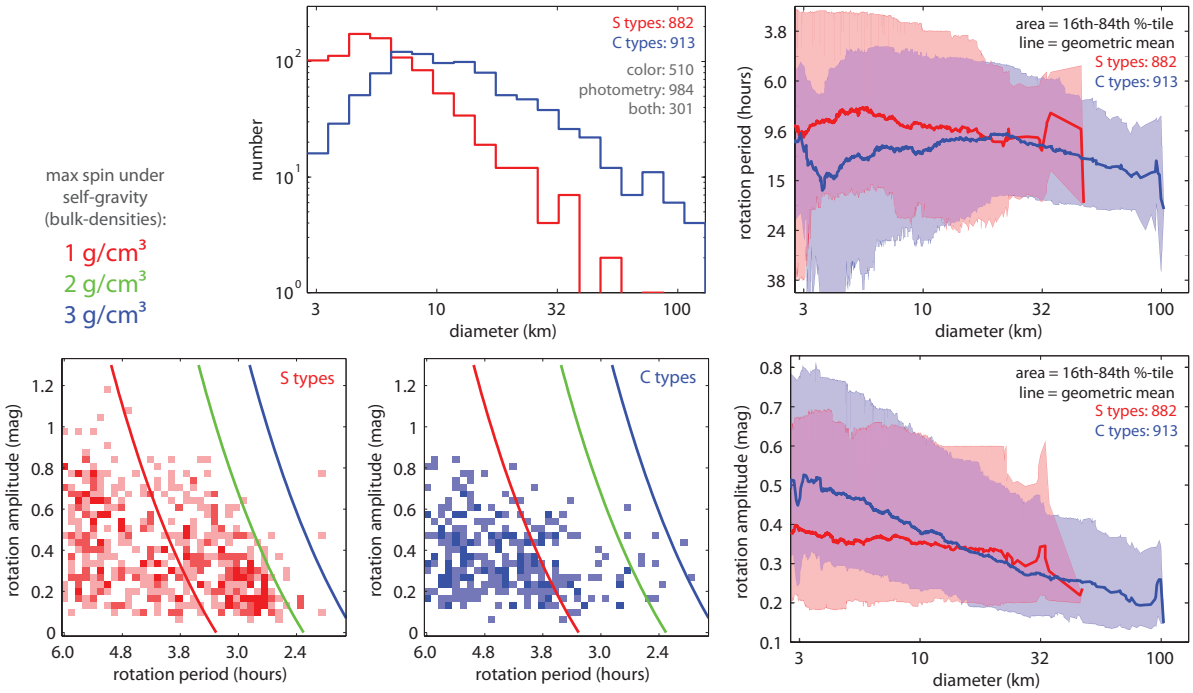


Figure 4.22: Taxonomic dependence on spin rate and amplitude, also versus diameter, using the union of the color-index and photometric-index based C/S taxonomy.

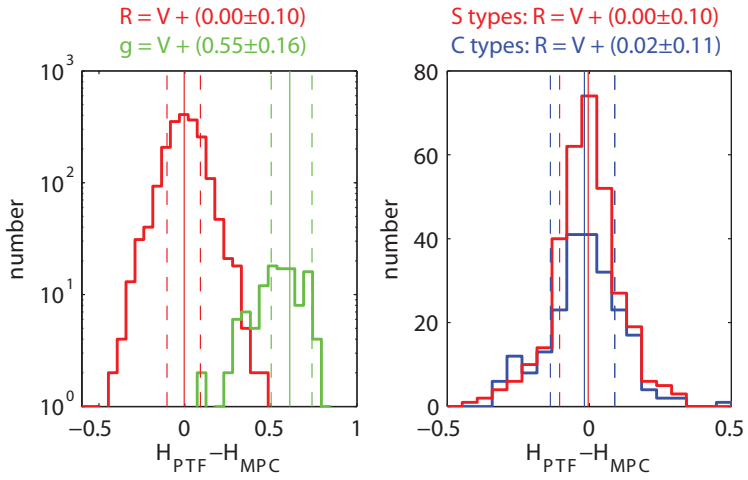


Figure 4.23: *Left*: Transformations between MPC V band and the PTF R and g bands for asteroids, based on the difference between MPC-fitted and PTF-fitted H magnitudes for asteroids whose PTF-fitted G values are in the range $0.10 < G < 0.20$ as well as other PTF-coverage constraints (see text). *Right*: R-band data only, with S and C types defined with either color and/or photometric indices (again using the <0.25 and >0.75 index thresholds).

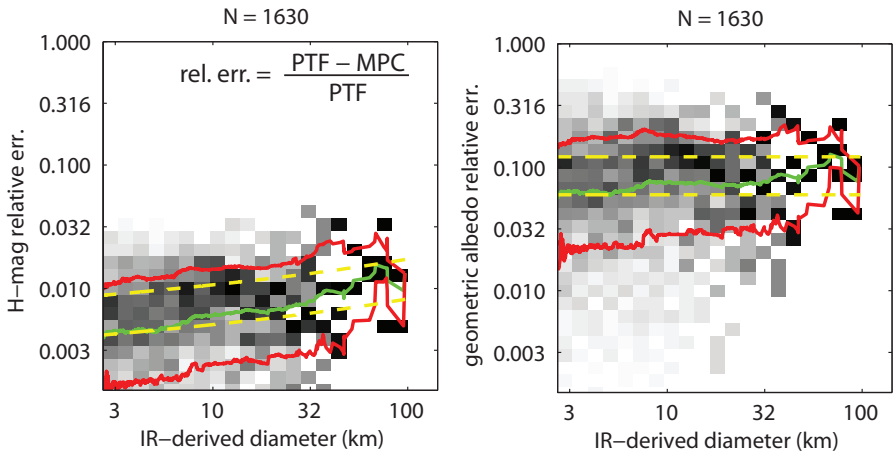


Figure 4.24: Errors in the MPC-listed absolute magnitudes relative to the PTF H values (in R band and using the G_{12} fit's H value), only considering asteroids with IR-derived diameters. On the right is the corresponding geometric albedo relative error. Pixels in the 2D histograms shown here are column normalized. The running-bin geometric mean and 16th and 84th percentiles are shown as green and red lines. Yellow dashed lines are the mean and 84th percentile expected from the 0.1 mag transformation uncertainty alone (for 7% geometric albedo).

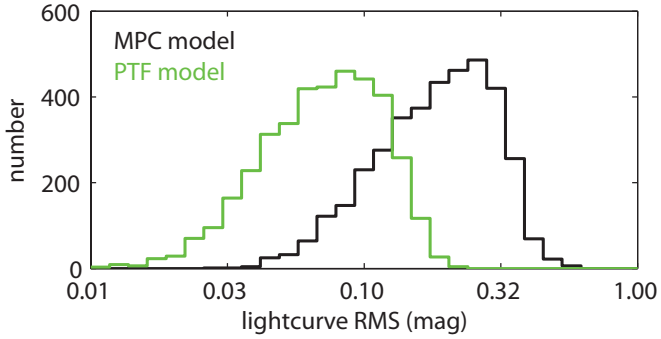


Figure 4.25: Comparison of the root-mean-square residuals, with respect to the PTF (H, G_{12}) plus rotation fit and the MPC (H, G) fit, for all lightcurves having a reliable R-band PTF phase-function fit.

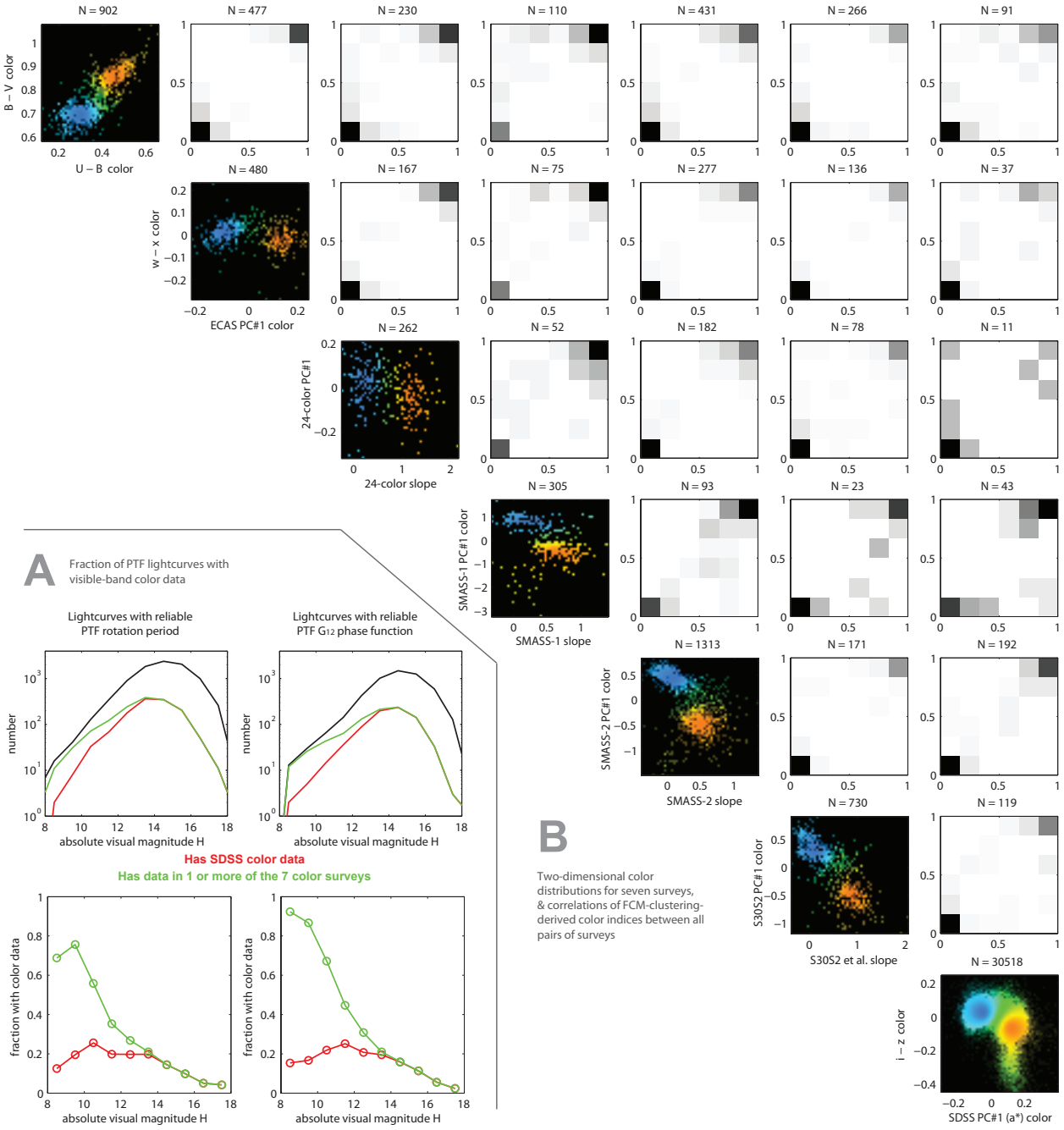


Figure 4.26: *Panel A*: Fraction of PTF lightcurves with colorimetric data available, for both the reliable-period and reliable-period-plus-G_{12} sets of lightcurves. *Panel B*: Two-dimensional color distributions for seven surveys, and correlations of FCM-clustering-derived classifications between all pairs of surveys.

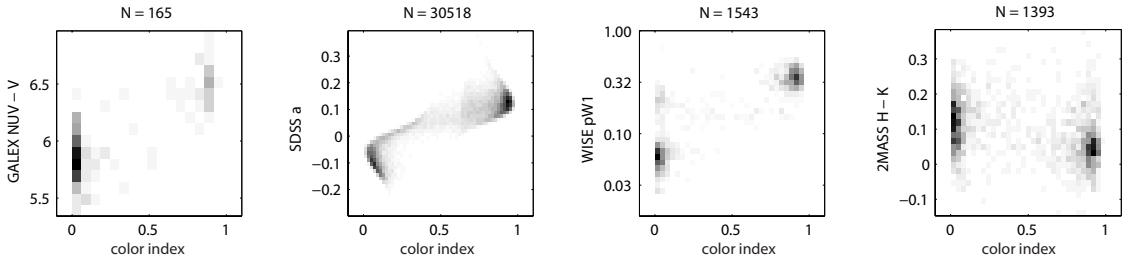


Figure 4.27: Relationship between various asteroid surface measurements (from the UV to near-IR) and this work's visible-color-derived C/S color index. See text for descriptions of data the data sets used here, and accompanying references.

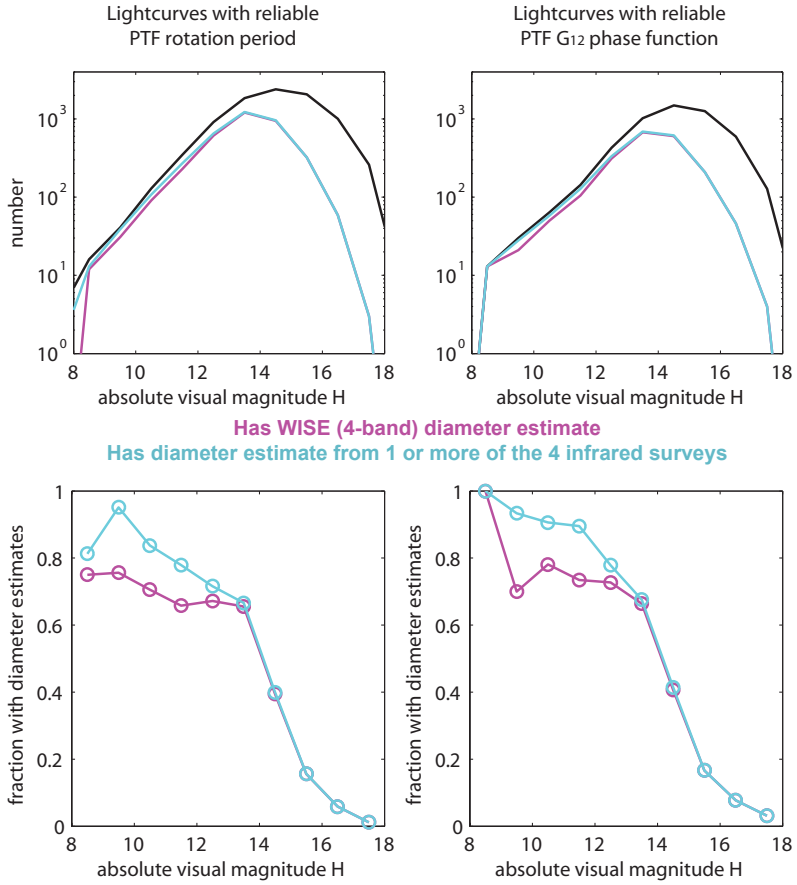


Figure 4.28: Fraction of PTF lightcurves with thermal-IR-based diameter estimates available, for both the reliable-period and reliable-G_{12} sets of lightcurves.

Chapter 5

Asteroids in *GALEX*

5.1 Introduction

As in visible wavelengths, ultraviolet flux from asteroids is entirely reflected sunlight. However, the steep drop in the solar spectrum shortward of ∼300 nm (Figure 5.1) makes asteroids orders of magnitude fainter in the UV than in the visible. For this reason—as well as the strong UV absorption by atmospheric ozone—UV observations of asteroids typically employ the *Hubble Space Telescope* (*HST*) or specialized instruments on a space-mission payload physically closer to the asteroid. These constraints have generally prohibited large-sample demographic studies of asteroids in the UV.

Predating *HST*, the *International Ultraviolet Explorer* (*IUE*) targeted 45 asteroids from 1978–1992, producing what remains to date the largest published sample of near-UV asteroid spectra [Roettger and Buratti (1994)], specifically in the range of 230–325 nm. The *IUE* data show evidence of clustering, principally with respect to geometric UV albedo. This clustering becomes further evident when coarsely-defined visible spectral type is included as a categorical parameter for each object (C, S and M types being the classes considered in the original work). Comparing the *IUE*-derived geometric UV albedos for each class with the geometric visible albedos demonstrated that the S types, which are redder-colored in the visible (specifically, 400–800 nm), remain redder than C types into the NUV. This is consistent with the observed spectral reflectance slope in this entire wavelength range (200–800 nm) being a broad absorption feature common to silica-bearing rocks (e.g., [Wagner et al. (1987)]).

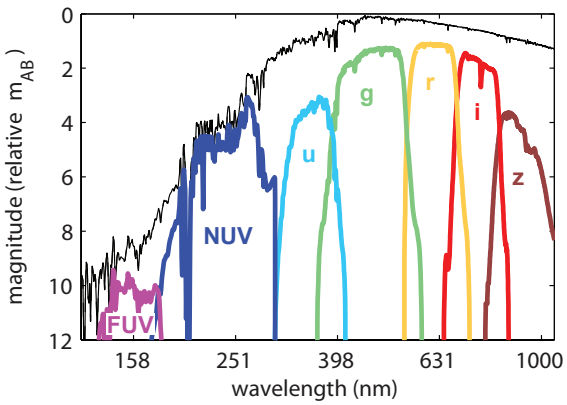


Figure 5.1: *GALEX* UV and SDSS visible filter response curves (colored lines) convolved with the spectrum of a G2 V type star (black line). The spectrum is from the library of [Pickles (1998)]. The vertical scale is in AB magnitude units per unit wavelength, offset by an arbitrary constant. Note wavelength is plotted on a log scale.

In this work we aim to verify this result with a new and larger sample of UV asteroid data from the *Galaxy Evolution Explorer* (*GALEX*), a NASA Small Explorer-class space telescope mission which from 2003–2012 conducted a UV imaging survey in a far-UV band (FUV, 130–190 nm) and a near-UV band (NUV, 180–280 nm). Approximately 2/3 of the sky was covered, with avoidance of bright stars and low galactic latitudes. [Martin et al. (2005))] discuss the extragalactic science program, while Morissey et al. (2005, 2007) discuss the on-orbit performance, survey calibration and data products. *GALEX* has a 50 cm^2 effective area, 1.25 degree diameter circular field of view, and FWHM resolution of 4.5″ in the NUV. Programs within the *GALEX* mission included an all-sky-survey (AIS, with ∼100 s exposures) and a medium-depth survey (MIS, with ∼1500 s exposures), and also a spectroscopic (grism) survey. Figure 5.1 shows the photometric response functions of the two *GALEX* bandpasses convolved with the solar spectrum, with comparison to the *ugriz* visible bandpasses. Detection of asteroids in the FUV is extremely unlikely (nonetheless, as described below we searched for both NUV and FUV asteroid detections).

Our approach in analyzing *GALEX* asteroid observations differs somewhat from Roettger and Buratti's treatment of the *IUE* data. First, instead of referencing taxonomic class labels (e.g., 'C type','S type','M type', etc.) assigned to individ-

Table 5.1: Observations of asteroids detected in *GALEX* NUV images. Includes 1,342 detections of 405 asteroids detected at least twice.

asteroid number	observation date (UT)	detected R.A. (deg)	detected Dec. (deg)	position residual ($''$)	NUV mag	NUV mag uncertainty	MPC-predicted visible mag ($V_{\rm MPC}$)	exposure time (s)	unique database ID (`objID` key in CasJobs)
1	2011-10-12.65797	355.22180	-18.47266	0.5	14.38	0.01	8.0	91	6380556162844065792
1	2011-10-21.42095	353.95617	-18.39556	0.2	14.67	0.01	8.1	80	6380556163951362048
3	2005-12-26.72536	75.12144	-1.31299	1.4	14.43	0.01	7.8	80	6381858059773280256
3	2011-04-17.04995	166.80140	8.52138	1.4	16.30	0.01	9.8	1513	3855329770719936512
6	2006-08-29.56208	309.25038	-19.31464	0.9	14.78	0.01	8.4	112	6379782093773209600
6	2005-05-07.46079	204.16985	11.58510	0.4	16.67	0.02	10.2	112	6378656257217134592
8	2004-12-21.41462	122.52225	19.17684	0.7	15.63	0.02	9.0	92	6377776615736213504
8	2004-12-21.48313	122.51009	19.18345	0.1	15.60	0.02	9.0	87	6377776615769767936

This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.

ual asteroids by previous authors, we define classes using a color index derived from a clustering analysis performed on a compilation of seven visible-color surveys [Waszczak et al. (2015)]. The brightest asteroids typically were targeted in one or more spectroscopic surveys—e.g., the Eight Color Asteroid Survey (ECAS, [Zellner et al. (1985)]) or the Small Main-Belt Asteroid Spectroscopic Surveys (SMASS; [Xu et al. (1995)], [Bus & Binzel (2002)]). Dimmer objects however often only have color information from the Sloan Digital sky-survey (SDSS; [York et al. (2000)], [Ivezić et al. (2001)], [Parker et al. (2008)]). The color index of [Waszczak et al. (2015)] puts asteroids of all sizes on a single, quantitative color scale (a proxy for spectral slope), the endmembers of which we identify with the C-type and S-type complexes. We use the terms 'C types' and 'S types' purely for compatibility with the literature, noting that our color index combined with the classification thresholds we apply to it represent original definitions of these two groups.

Use of a one-dimensional color metric sacrifices sensitivity to secondary taxonomic groups such as M types, D types and V types. While our purpose is mainly to examine the *first-order* taxonomic variation, in certain cases we divide our color-defined 'C types' into two classes (C_{high} and C_{low}) on the basis of visible and near-infrared albedo (as detailed in Section 5.3.2). At the end of this work we also examine well-established C-complex subgroups from both the Tholen and Bus/Binzel taxonomies (Section 5.6), showing how these known subgroups (e.g., M types and G types) are distributed in this work's color-albedo space and highlighting these subgroups' unique UV properties.

A second distinction between this work's approach and [Roettger and Buratti (1994)] is that, rather than comparing the *geometric albedo* in the UV with that of the visible band, we focus on the difference in *apparent magnitudes* between the UV and visible. One motivation for doing this is we need not make any assumptions about the phase function of asteroids in the UV. The challenge however is that we must accurately estimate the visible flux at the time of the UV observations. As discussed in Section 5.3, we adopt (and compare) two distinct methods for predicting the visual

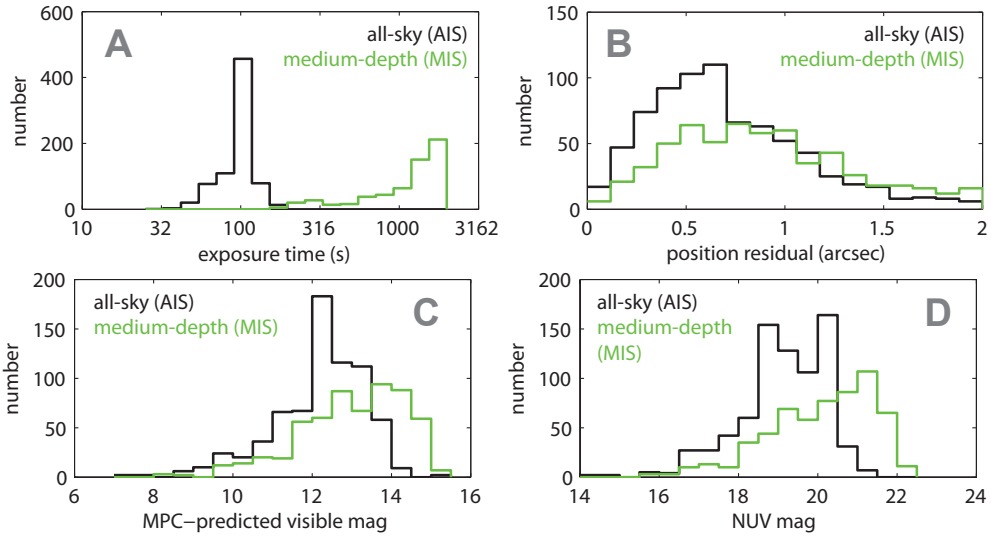


Figure 5.2: Characteristics of positive asteroid detections from the two *GALEX* surveys (distinguished by exposure time) shown separately in black and green.

magnitude. The first method simply adopts the widely-used MPC[1] predicted magnitudes; the second method applies color-dependent phase-function and bond-albedo estimates adapted from the [Waszczak et al. (2015)] study of lightcurves from the Palomar Transient Factory survey[2] (PTF; [Law et al. (2009)]; [Rau et al. (2009)]).

5.2 GALEX asteroid observations

Extracting detections of known asteroids from a survey involves a three-dimensional (R.A., Dec., time) cross-matching of the ephemeris of all asteroids against the survey's time-stamped image boundaries. We modified software originally used to search for asteroids in PTF ([Waszczak et al. (2013b)], [Waszczak et al. (2015)]) to instead search for asteroids in *GALEX*.

We first retrieved the metadata of all *GALEX* images, available from the Space Telescope Science Institute via command-line queries with the CasJobs tool [Li and Thakar (2008)]. We then indexed all image centers with respect to (R.A, Dec.) into uniformly-spaced sky cells of 3-degree radius. For all ∼380,000 numbered

[1]IAU Minor Planet Center, http://minorplanetcenter.net
[2]http://ptf.caltech.edu

asteroids, we queried JPL's online service HORIZONS [Giorgini et al. (1996)] to generate a 1-day-spaced ephemeris spanning 2003–2012. Using an object-specific search radius equal to 3 degrees (cell radius) plus 0.75 degrees (FOV radius) plus the object's maximum 1-day motion ($\sim$10 arcminutes for most main-belt objects), we matched the ephemeris points against the sky cells. For each matched cell, we filtered out all images in that cell not within the epoch range of the matched ephemeris points, then for each surviving image we re-queried HORIZONS for the precise location at each observed epoch. We next performed a 1.25-degree-radial match of these precise positions against the relevant *GALEX* image centers.

We found $\sim$850,000 predicted detections of numbered asteroids (with no limit on apparent magnitude) in *GALEX* using this method. For each predicted detection, using CasJobs we queried the *GALEX* single-visit source list (as opposed to the co-added source list). Multiple matches near the same point occurring more than 6 hours apart were excluded, as were all matches further than $2''$ from the predicted location. Additionally, to ensure the inclusion of greater than (approximately) 5σ detections, we discarded all matches with NUV > 21 mag in the shorter exposures (AIS program), and discarded all matches with NUV > 22.7 mag in the longer exposures (MIS program), following the limiting magnitudes quoted by [Morissey et al. (2007))].

Following the above procedure and criteria, we extracted a total of 1,342 positive NUV detections of 405 unique asteroids which were detected by *GALEX* at least twice (and no FUV detections, as expected). These detections are listed in Table 5.1; several histograms detailing these detections appear in Figure 5.2.

5.3 Modeling visible magnitudes

In this section we consider two distinct methods of estimating the visible magnitudes corresponding to all *GALEX* NUV detections; this in turn provides the distribution of the asteroids' NUV $- V$ color. The general model for an asteroid's apparent visual magnitude V (log flux) is

$$V = H + \delta + 5\log_{10}(r\Delta) - 2.5\log_{10}[\phi(\alpha)], \tag{5.1}$$

where H is the absolute magnitude (a constant), δ is a periodic variability term due to rotation (e.g., if the object is spinning and has some asymmetry in shape or albedo), r and Δ are the heliocentric and geocentric distances (in AU), and $\phi = \phi(\alpha)$ is the *phase function*, which varies with the solar phase angle α (the Sun-asteroid-Earth angle). When $\alpha = 0$ (i.e., at opposition), $\phi = 1$ by definition, while in general $0 < \phi < 1$ for $\alpha > 0$ (with ϕ decreasing as α increases).

All asteroids for which we have extracted *GALEX* observations have known orbits, meaning r, Δ, and α are accurately and precisely known at all observed epochs. Our two methods for estimating V differ in their assumptions regarding (and observational data used to constrain) H and ϕ. In both cases we do not attempt to model the rotational term δ, but rather incorporate δ into the uncertainty of V using lightcurve amplitude estimates from the literature. In particular, 388 of the 405 *GALEX*-observed asteroids have an amplitude lower-limit estimate available in the Lightcurve Database ([Warner et al. (2009)], [Harris et al. (2012)]).

In the following sections we refer to two different albedo quantities. The visible-band *geometric* albedo p_V relates to the visible-band *bond* albedo $A_{\rm bond}$ and the phase function ϕ (of Equation [5.1]) according to

$$p_V \equiv \frac{A_{\rm bond}}{2}\left(\int_0^\pi \phi(\alpha)\sin(\alpha)\,d\alpha\right)^{-1} \equiv \frac{A_{\rm bond}}{q}, \tag{5.2}$$

The above equation also defines the phase integral q. The bond albedo $A_{\rm bond}$ is defined as the total visible light energy reflected or scattered by the asteroid (in all directions) divided by the total visible light energy incident upon the asteroid (from the Sun). Assuming the asteroid has a circular cross-section of diameter D, this can be expressed as

$$A_{\rm bond} \equiv \frac{\int_0^\pi f(\alpha)\sin(\alpha)d\alpha}{(f_{\rm Sun}/4\pi{\rm AU}^2)\times\pi(D/2)^2}, \tag{5.3}$$

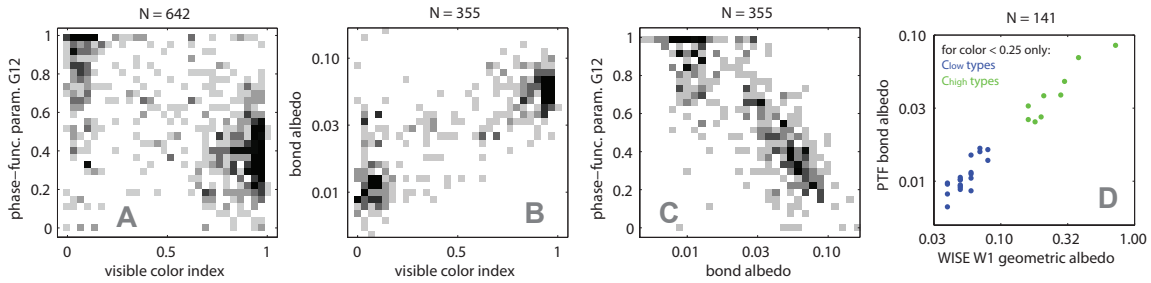


Figure 5.3: We compute $V_{\rm PTF}$ model magnitudes by first assigning fixed $A_{\rm bond}$ and G_{12} values to each *GALEX*-observed asteroid depending on its color class; we then use D to compute H, and finally use the assumed G_{12} value to predict V. The fixed values of $A_{\rm bond}$ and G_{12} are medians from the color-albedo-G_{12} data in [Waszczak et al. (2015)], 2D histograms of which are shown here. Above each plot is the sample size ($N = \ldots$). A total of 642 asteroids have color data *and* G_{12} values in the PTF data; 355 of these also have diameters available (required to compute $A_{\rm bond}$). Panel D shows that *WISE* $W1$ geometric albedos correlate with the PTF bond albedo; we thus use the *WISE* p_{W1} data to assign C types either a low ($A_{\rm bond} \approx 0.01$) or high ($A_{\rm bond} \approx 0.04$) bond albedo.

where $f(\alpha) = 10^{-V(\alpha)/2.5}$ is the asteroid's flux as a function of phase angle, with $V(\alpha) = H - 2.5\log_{10}\phi(\alpha)$ being Equation (5.1) evaluated at $\delta = 0$ and $r = \Delta = 1$ AU (similarly, $f_{\rm Sun} = 10^{-V_{\rm Sun}/2.5}$).

5.3.1 (H, G) from MPC data

The first method for estimating V adopts the Minor Planet Center's computed absolute magnitudes ($H_{\rm MPC}$), which are regularly updated by the MPC's automated processes and utilize the Lumme-Bowell G-parameter model for ϕ [Bowell et al. (1989)]. This same (H, G) model then predicts the apparent magnitude $V_{\rm MPC}$ as a function of solar phase angle.

The $H_{\rm MPC}$ values are fit to photometry provided by a variety of surveys/individuals, many of whom may use slightly different absolute calibration standards or filters with slightly different specifications. A small fraction of asteroids have fitted G values; [Harris & Young (1988)] present mean G values for several major taxonomic classes, with $G = 0.15$ being an average between the C types ($G \approx 0.08$) and the S types ($G \approx 0.23$). For the majority of asteroids the MPC uses an assumed $G = 0.15$ with

this model. [Waszczak et al. (2015)] compares the $H_{\rm MPC}$ values with H magnitudes derived from a model that includes rotation and the more modern (H,G_{12}) phase function of [Muinonen et al. (2010))]. Among bright asteroids the relative difference is typically between 0.3% to 3%, corresponding to (on average) an $\sim$0.07 mag discrepancy.

Though $H_{\rm MPC}$ values are available for all 405 *GALEX*-observed asteroids, we only consider the subset of 315 asteroids having visible-band color indices of either less than 0.25 ('C types') or greater than 0.75 ('S types'). Of these, 41 asteroids have $G_{\rm MPC} \neq 0.15$.

5.3.2 (D,$A_{\rm bond}$,G_{12}) from PTF, infrared, and color data

Our second means of estimating visual magnitudes applies only to asteroids having both a color index *and* a diameter estimate constrained from thermal fluxes in an infrared survey[3]. In this approach we use the G_{12}-parameter model for ϕ [Muinonen et al. (2010))], and we replace H with its equivalent expression[4] in terms of the diameter D, bond albedo $A_{\rm bond}$, and phase integral q:

$$H = -5 \log_{10} \left(\frac{D\sqrt{A_{\rm bond}/q}}{1329 \text{ km}} \right), \tag{5.4}$$

where the phase integral q is a linear function of G_{12}:

$$q(G_{12}) = \begin{cases} 0.2707 - 0.236 G_{12} & \text{if } G_{12} < 0.2; \\ 0.2344 - 0.054 G_{12} & \text{otherwise.} \end{cases} \tag{5.5}$$

We again define 'C types' as all asteroids with color indices less than 0.25 and 'S types' as all with color indices greater than 0.75. For S types we then consider

[3]Similar to the color data, the diameter data set we use is a compilation of products from several surveys and described in the appendix of [Waszczak et al. (2015)]. The source IR surveys are *WISE* ([Wright et al. (2010)], [Masiero et al. (2011)], [Masiero et al. (2014)]), *IRAS* ([Matson et al. (1986))], [Tedesco et al. (2002a)]), *MSX* [Tedesco et al. (2002b)], and AKARI [Usui et al. (2011)].

[4]Equation (5.4) follows *directly* from combining Equations (1)–(3). The constant 1329 km depends on somewhat arbitrarily-defined quantities such as the Sun's visual magnitude and the ratio of an AU to a kilometer.

Table 5.2: $A_{\rm bond}$ and G_{12} (based on PTF data) of color-defined taxonomic groups.

class name	color index	*WISE* p_{W1}	$A_{\rm bond}$ median	$A_{\rm bond}$ scatter*	G_{12} median	G_{12} scatter*
S	> 0.75	N.A.	0.056	0.016	0.36	0.16
$C_{\rm high}$	< 0.25	> 0.125	0.038	0.022	0.42	0.20
$C_{\rm low}$	< 0.25	< 0.125	0.010	0.003	0.84	0.16

*Scatter is here defined as $0.5 \times (84^{\rm th}$ percentile $-16^{\rm th}$ percentile)

diameters derived from any of four infrared surveys (see Footnote 3), while for C types we specifically require that the asteroid have been observed in the *WISE* 4-band cryogenic survey ([Wright et al. (2010)], [Masiero et al. (2014)] and references therein). Both the *WISE* $W1$-band geometric albedo p_{W1} and the PTF-derived bond albedo[5] $A_{\rm bond}$ show evidence of bimodality among objects with color indices less than 0.25 (Figure 5.3 panel D). Thus, we divide the C types into low-bond-albedo ($C_{\rm low}$) and high-bond-albedo ($C_{\rm high}$) subgroups based on their p_{W1} as reported by [Masiero et al. (2014)]. In Section 5.6 we show that the $C_{\rm high}$ types most closely correspond to what other authors have called M types.

[Waszczak et al. (2015)] computed $A_{\rm bond}$ and G_{12} values for $\sim$1,600 asteroid lightcurves in the PTF survey. Using that work's data (Figure 5.3) we compute median $A_{\rm bond}$ and G_{12} values (and associated scatter) for the S, $C_{\rm low}$ and $C_{\rm high}$ taxonomic groups. Table 5.2 summarizes the definitions and assumed $A_{\rm bond}$ and G_{12} values of these groups. There are 245 *GALEX*-observed asteroids (out of the 405 in Table 5.1) which have color *and* diameter data available, allowing them to be modeled by this method. To each *GALEX*-observed asteroid we assign the appropriate $A_{\rm bond}$ and G_{12} value based on its class membership, then use its diameter to compute a model absolute magnitude ($H_{\rm PTF}$) using Equation (5.4). Together with the assumed G_{12} value, this $H_{\rm PTF}$ then predicts the apparent magnitude $V_{\rm PTF}$ at each *GALEX*-observed solar phase angle.

[5]The visible bond albedo $A_{\rm bond}$ uses the same *WISE* diameter used by [Masiero et al. (2014)] in computing the $W1$-band geometric albedo p_{W1}.

5.3.3 Rotational uncertainty in V

Both the $V_{\rm MPC}$ and $V_{\rm PTF}$ model magnitudes discussed here lack an estimate of the rotational term (δ in Equation [5.1]). We account for this by incorporating a term for rotational modulation into the reported uncertainty of V. Of the 315 asteroids with $V_{\rm MPC}$ values, 302 have an amplitude lower limit listed in the Lightcurve Database ([Warner et al. (2009)], [Harris et al. (2012)]) , while for the 245 asteroids with $V_{\rm PTF}$ predictions there are 239 with reported amplitudes. As shown for instance by [Waszczak et al. (2015)], asteroids in the relevant size range typically have amplitudes less than ∼0.4 mag. For the few objects in our sample lacking an amplitude limit, we assume a value of 0.2 mag.

Assuming an asteroid's rotational phase φ to be random at the time of a *GALEX* detection (i.e., with a probability distribution of the form $P(\varphi) \propto$ constant), then the probability distribution of a basic sinusoidal δ (i.e., one of the form $\delta = \delta_0 \sin\varphi$) can be shown to have the form

$$P(\delta) \propto \frac{1}{\sqrt{\delta_0^2 - \delta^2}}, \tag{5.6}$$

where δ_0 is the amplitude. We use Equation (5.6) as a probability density function to generate, for each modeled V, a set of 10^4 simulated δ values. These simulated δ are added to an equal number of model V magnitudes computed by random (Gaussian distribution) sampling of the component terms: in the case of $V_{\rm MPC}$ we just assume a fixed $H_{\rm MPC}$ uncertainty of 0.1 mag, whereas for the $V_{\rm PTF}$ values we randomly sample all three of $A_{\rm bond}$, G_{12}, and D, using the scatter values in Table 5.2 for the first two and the literature-reported diameter uncertainty for D. The $16^{\rm th}$ to $84^{\rm th}$ percentile spread in the distribution of combined $\delta + V$ values then becomes the quoted uncertainty for V.

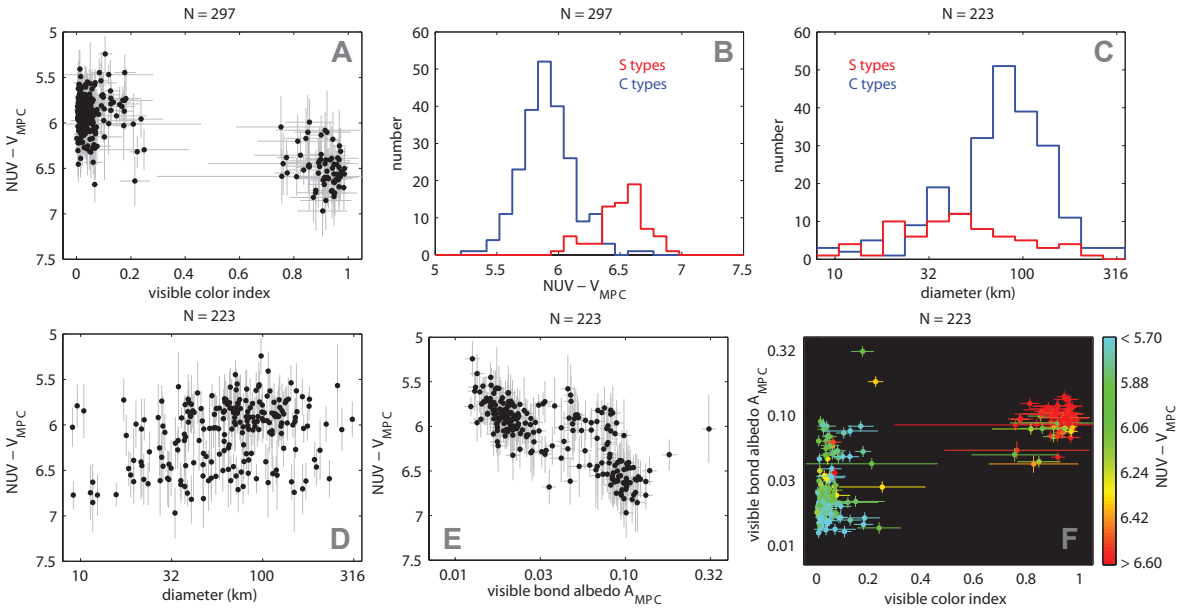


Figure 5.4: Distribution of the NUV $-$ V color for *GALEX*-observed asteroids using the H, G model with MPC data to predict V. Plots B and C define C types and S types as objects with color indices of <0.25 and >0.75, respectively. Plots C–F include only the subset with a diameter estimate available; this subset is precisely the same sample considered in Figure 5.6.

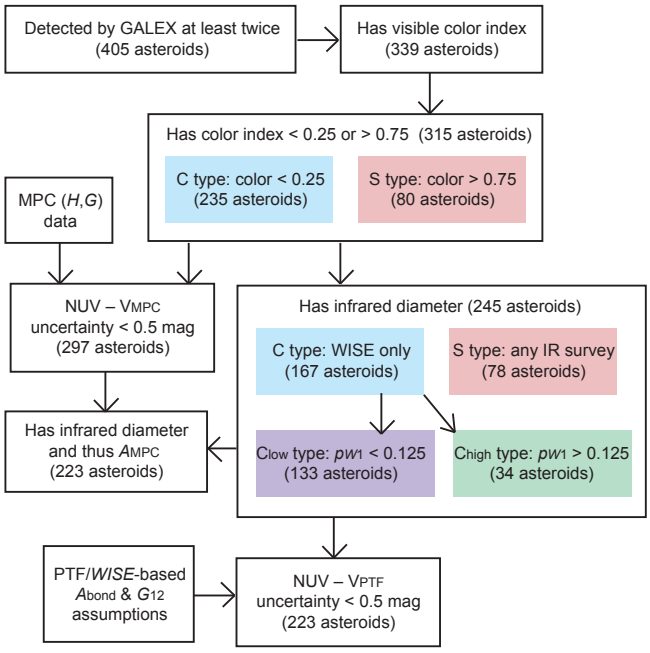


Figure 5.5: Flowchart visualizing the steps in the *GALEX*-observed asteroid sample selection process. Each box is a subset of the box pointing to it.

Table 5.3: NUV $-$ V color (mag. units) of *GALEX*-observed asteroids and sample sizes

class name	NUV $-$ $V_{\rm MPC}$ median	NUV $-$ $V_{\rm MPC}$ scatter*	NUV $-$ $V_{\rm PTF}$ median	NUV $-$ $V_{\rm PTF}$ scatter*	$N_{\rm MPC}$	$N_{\rm PTF}$
S	6.52	0.25	6.71	0.21	72	69
C	5.90	0.19	6.03	0.22	225	154
$C_{\rm high}$	–	–	6.14	0.33	–	29
$C_{\rm low}$	–	–	6.02	0.19	–	125

*Scatter is here defined as $0.5 \times$ ($84^{\rm th}$ percentile $-16^{\rm th}$ percentile)

5.4 NUV $-$ V color distribution

Having computed the model V magnitudes, we obtain the NUV $-$ V color for each *GALEX* asteroid detection and the corresponding uncertainty. The latter contains an additional rotational uncertainty component (now associated with the NUV observation), again determined by repeated sampling of Equation (5.4) as described above. Since all the asteroids we consider have more than one *GALEX* NUV detection, we compute the variance-weighted average NUV $-$ V color for each asteroid (plotted in Figures 4 and 6); the uncertainty in this average is the inverse quadrature sum of the individual uncertainties.

In Figures 4 and 6 (and the accompanying analysis) we have omitted all asteroids with NUV $-$ V uncertainties of greater than 0.5 mag. As a result, the sample size of asteroids with NUV $-$ $V_{\rm MPC}$ estimates is 297 (out of the 315 quoted in Section 5.3.1), while the sample with NUV $-$ $V_{\rm PTF}$ estimates is 223 (out of the 245 quoted in Section 5.3.2). Figure 5.5 graphically summarizes the sample selection criteria in a flowchart. In Figures 4, 6 and 7, the errorbars on the color indices were computed by a bootstrapping process described in the appendix of [Waszczak et al. (2015)].

Both the $V_{\rm MPC}$ and $V_{\rm PTF}$ model magnitudes produce a bimodal NUV $-$ V color distribution, with the S types having the redder NUV $-$ V color (panels A and B of both Figures 4 and 6). Median and scatter of NUV $-$ V for the various classes appear in Table 5.3. To formally ascertain the inequality of the two distributions, we use the two-sided Kolmogorov-Smirnov (KS) test [Massey (1951))], which compares two

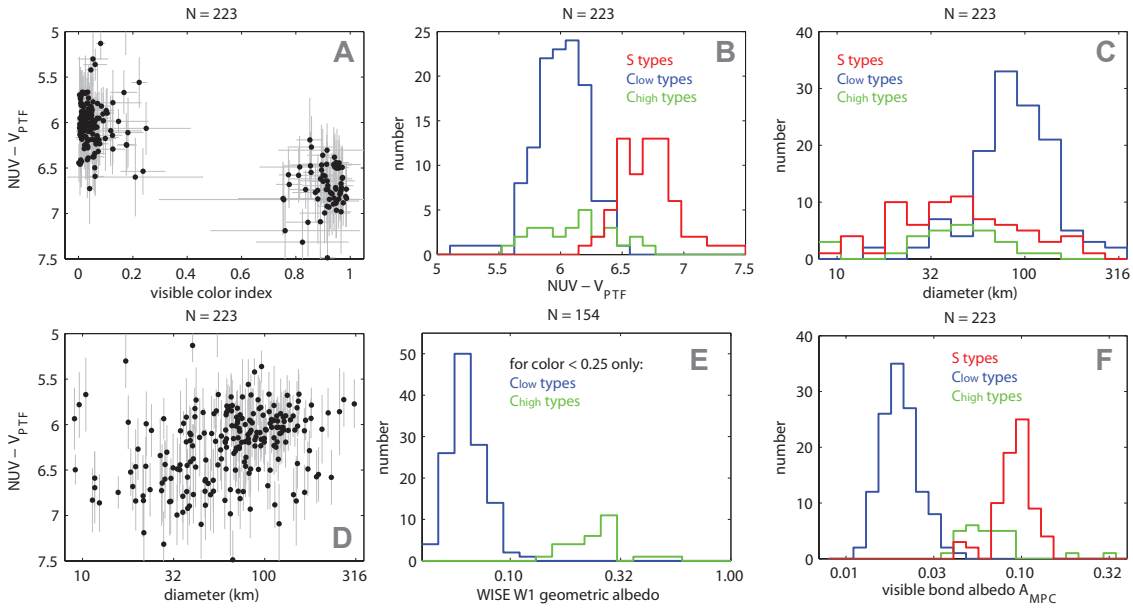


Figure 5.6: Distribution of the NUV $-$ V color for *GALEX*-observed asteroids using the D,$A_{\rm bond}$,G_{12} model with PTF data, infrared data, and color data to to predict V. See Table 5.2 for the definitions of the S, $\rm C_{low}$, and $\rm C_{high}$ groups in plots B, C, E, and F.

empirical distributions via a bootstrap method. In particular this test computes a statistic quantifying the extent to which the cumulative distribution function differs in the two distributions being compared. For the $V_{\rm MPC}$ model we find the C-type NUV $-$ V color distribution differs from that of the S-type distribution at an 11.6σ significance level (Figure 5.4 panel B). For the $V_{\rm PTF}$ model (Figure 5.6 panel B) we find the C types ($\rm C_{low}$ and $\rm C_{high}$ combined) differ from the S types at an 8.1σ level, while the $\rm C_{low}$ and $\rm C_{high}$ types only differ at a 1.9σ level (this difference is thus not statistically significant).

An important characteristic of our sample is that the C types outnumber the S types by a ratio of 3:1 in the $V_{\rm MPC}$ sample and a ratio of 2:1 in the $V_{\rm PTF}$ sample (cf. panel C of Figures 4 and 6). This ratio indirectly reflects both an inherent difference in the populations of the two types as well as a detection bias due to S types typically having smaller orbital distances and thus typically brighter apparent magnitudes for a given size.

In panels C–F of Figure 5.4 the sample size decreases from $N = 297$ down to $N = 223$ asteroids as we consider only those objects in the $V_{\rm MPC}$ sample that also have available diameters (this is equivalently the $V_{\rm PTF}$ sample considered in Figure 5.6). We compute the MPC-based visible bond albedo $A_{\rm MPC}$ using Equation (5.2) together with the asteroid's $H_{\rm MPC}$ and $G_{\rm MPC}$ values. In particular, there are 38 asteroids (out of the 223 with diameters) with a measured $G_{\rm MPC} \neq 0.15$; for the remainder we assume $G_{\rm MPC} = 0.15$ for consistency with the manner in which the $V_{\rm MPC}$ are computed. Analogous to Equation (5.5), the phase integral for the G-model (required for computation of $A_{\rm MPC}$ via Equation [5.4]) is

$$q(G) = 0.290 + 0.684G, \tag{5.7}$$

as given by [Bowell et al. (1989)]. With the $G_{\rm MPC} = 0.15$ assumption for the majority of the asteroids in our sample, the $A_{\rm MPC}$ values are not expected to be as accurate as the $A_{\rm bond}$ values computed for instance by [Waszczak et al. (2015)], wherein distinct q values were fitted to each object on the basis of a lightcurve. Nonetheless, it is instructive to compute $A_{\rm MPC}$, e.g., to check for consistency with the class-median $A_{\rm bond}$ values, and to exploit as a second taxonomic metric in addition to visible color.

Figure 5.4 panel E shows that NUV$-V$ correlates with $A_{\rm MPC}$ ($\rho_{\rm Spearman} = 0.698$, $>10\sigma$ significance), similar to how NUV$-V$ correlates with the color index in panel A ($\rho_{\rm Spearman} = 0.491$, $>10\sigma$ significance). Unlike the color index however, the separation between the $\rm C_{low}$ and $\rm C_{high}$ subgroups is qualitatively evident in this plot. Figure 5.4 panel F combines all three parameters; note the axes are the same as Figure 5.3 panel B, with $A_{\rm MPC}$ replacing $A_{\rm bond}$ and the data consisting of *GALEX*-observed asteroids rather than PTF-observed asteroids.

Figure 5.6 panel F confirms (independently of Figure 5.3 panel D) the validity of using *WISE* $W1$-band geometric albedo as a proxy for visible bond albedo to separate $\rm C_{low}$ from $\rm C_{high}$—the two classes robustly differ in their $A_{\rm MPC}$ distributions (9.5σ KS-test significance). However, the class-median $A_{\rm MPC}$ values of the $\rm C_{low}$, $\rm C_{high}$, and S types are 100%, 67%, and 63% *greater* than their class-median PTF-based $A_{\rm bond}$

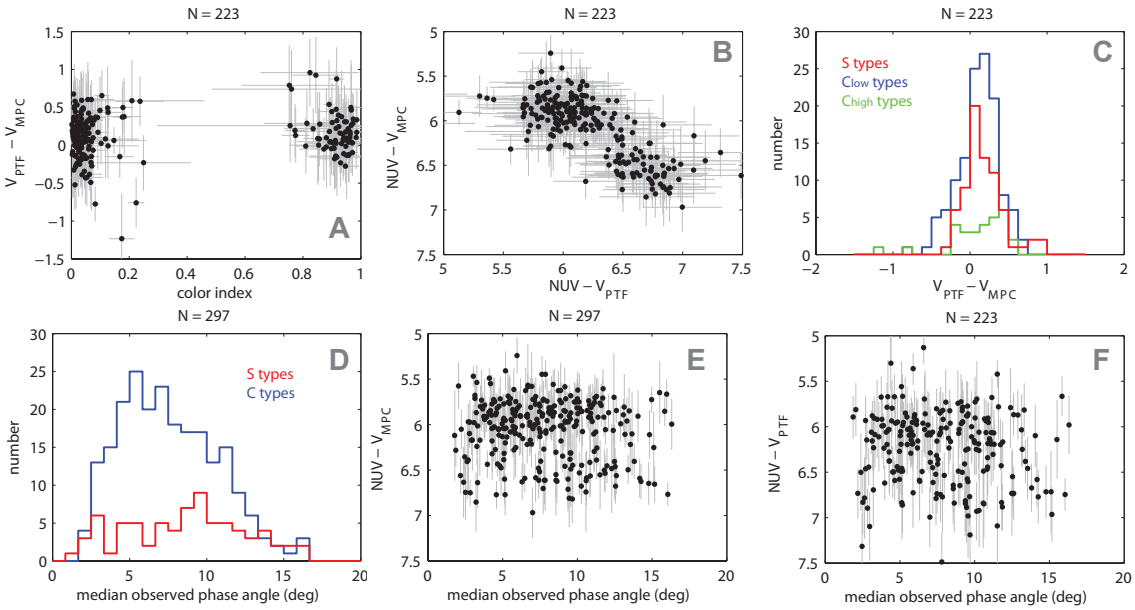


Figure 5.7: Various checks for systematic differences in the predicted V magnitudes output by the two different photometric models. *Bottom row*: Investigation of phase-angle-dependence on the NUV $-$ V color.

values in Table 5.2. This reflects the differing values of H and q produced by the G and G_{12} models, as well as the fact that we apply class-specific G_{12} values, whereas $G_{\rm MPC} = 0.15$ is assumed for the majority of asteroids, regardless of their class.

Consideration of both the $V_{\rm MPC}$ and $V_{\rm PTF}$ model magnitudes provides two independent means of computing NUV $-$ V; this helps rule out the effect of potential systematic errors unique to either one of the V models, as well as possible biases in the distinct observational data sets upon which each V is based. In Figure 5.7 panels A–C we examine the distribution of $V_{\rm MPC} - V_{\rm PTF}$ for all 223 asteroids having both V estimates. The median of $V_{\rm MPC} - V_{\rm PTF}$ is 0.13 mag (scatter of 0.25 mag), indicating the MPC-based model consistently produces brighter V estimates. For C types the median $V_{\rm MPC} - V_{\rm PTF}$ is 0.14 mag and for S types it is 0.12 mag; the two groups' $V_{\rm MPC} - V_{\rm PTF}$ distributions differ with less than 0.1σ significance in a KS-test.

In Section 5.1 we motivated our choice to examine the difference in *apparent magnitude* between UV and visible (as opposed to the difference in *albedo* in UV and visible) by noting that little is known of asteroid phase functions in the UV,

rendering difficult the estimation of UV absolute magnitudes (and hence UV albedos). A potential issue with this approach which we have heretofore ignored is that, if the phase function *does* differ significantly in the UV from the visible, then the NUV $-V$ color will vary with phase angle. Figure 5.7 panels D–F attempt to ascertain whether such a trend exists by considering the median phase angle at which each asteroid was detected by *GALEX*.

C types are observed at a median median phase angle of 7.2 deg compared to the S types' median median phase angle of 9.0 deg. This is explained by the fact that C types on average have larger semi-major axes, which geometrically correspond to lower observed phase angles from Earth. Within the C-type group, median phase angle correlates with NUV $- V_{\mathrm{MPC}}$ at $\rho_{\mathrm{Spearman}} = -0.1$ (1.5σ significance) and with NUV $- V_{\mathrm{PTF}}$ at $\rho_{\mathrm{Spearman}} = 0.01$ (0.1σ significance). Among S types, median phase angle correlates with NUV $- V_{\mathrm{MPC}}$ at $\rho_{\mathrm{Spearman}} = 0.01$ (0.1σ significance) and with NUV $- V_{\mathrm{PTF}}$ at $\rho_{\mathrm{Spearman}} = 0.07$ (0.5σ significance). We therefore cannot claim any phase angle dependence for NUV $- V$, regardless of the taxonomic group or V-model being considered.

Various works (e.g., [Sanchez et al. (2012)] and references therein) discuss the phenomenon of asteroid phase reddening, i.e. an observed reddening of visible color with increasing phase angle. Very few survey-scale samples have been used to test for the presence of this effect. [Szabó et al. (2007)] computed slightly different phase-angle dependences for the $g - r$ and $r - i$ colors of Trojans in SDSS, though these relations were not separately computed for the Trojans' two taxonomic groups. [Waszczak et al. (2015)] did not detect any statistically significant difference between G_{12} fits to r-band PTF lightcurves and g-band PTF lightcurves (among asteroids that had data in both bands). The extent to which a phase-function dependence on wavelength exists between the UV and visible remains unclear. Future UV surveys such as *ULTRASAT* [Sagiv et al. (2014)] offer the most promising means of testing this hypothesis, especially because (unlike *GALEX*) they will obtain sufficient numbers of observations to adequately sample UV lightcurves, thereby providing the best possible data set for fitting UV phase functions.

5.5 Albedo vs. wavelength

If we assume that the phase function does not differ significantly between the UV and visible (or take Figure 5.7 panels E and F as justification of this statement), then we can compare the relative *bond albedo* versus wavelength for the different taxonomic groups using measured colors, filter response functions and the solar spectrum. If the phase function *does* in fact vary significantly with wavelength, then this approach only provides the relative *geometric albedo* versus wavelength (see Equations [2] and [3]).

Assume photometry from two filters (1 and 2) produce the color measurement $m_1 - m_2$. This color relates to the solar flux distribution $S(\lambda)$, the albedos in each band (A_1 and A_2) and the filter responses $F_1(\lambda)$ and $F_2(\lambda)$ according to

$$10^{(m_1-m_2)/2.5} = \frac{\int F_1(\lambda)\lambda^{-2}d\lambda \int S(\lambda)F_2(\lambda)A_2 d\lambda}{\int S(\lambda)F_1(\lambda)A_1 d\lambda \int F_2(\lambda)\lambda^{-2}d\lambda} \tag{5.8}$$

which we adapted from a similar equation in [Pickles (1998)].

Using NUV as band 1 and V as band 2 in Equation 5.6, we use the colors in Table 5.3 (specifically, the $V_{\rm PTF}$-based colors) to obtain the albedo ratio $A_{\rm NUV}/A_V$, with uncertainties coming from the associated scatter in the colors. In Figure 5.7 we plot these albedo ratios for the C types and S types, incorporating an additional uncertainty component from the transformation from r to V (see [Waszczak et al. (2015)] for a discussion of this transformation in the context of asteroids). The end-computed values are $(A_{\rm NUV}/A_r)_{\rm C} = 0.63^{+0.14}_{-0.12}$ and $(A_{\rm NUV}/A_r)_{\rm S} = 0.33^{+0.07}_{-0.06}$. The relative albedo values in the SDSS bands included for comparison in Figure 5.8 are taken directly from a figure in [Ivezić et al. (2001)]; likewise the ECAS data are taken directly from a figure in [Zellner et al. (1985)].

Note that Figures 8 and 9 ignore the fact that the S types' and C types' *absolute albedo* in r-band differs. In other words, these plot could be converted into ones with absolute albedo on the vertical scale by multiplying the blue and red lines by their respective absolute r-band albedos, which would be similar to those listed in Table 5.2 for V band.

Both C and S types show a continued trend of decreasing albedo at shorter wavelengths. Whereas for S types this behavior was already well-established in the 300–800 nm region, for C types the $u - g$ color had previously represented a significant deviation from the shallower slope observed from 400–800 nm. The C-type NUV albedo in Figure 5.8 confirms the presence of a marked drop in albedo somewhere in the 200–400 nm range. Given the resolution of Figure 5.8 and the uncertainties in the data points, we cannot judge whether the C-type albedo levels off between NUV and u band, or whether the slope between u and g bands persists into these shorter wavelengths. The *IUE* spectra from [Roettger and Buratti (1994)] indicate C-type albedo is constant at least in the range 240–300 nm (as does the Lutetia data described below), so that the former interpretation may be more accurate.

5.6 Comparison to *HST* data

5.6.1 Lutetia

[Weaver et al. (2010)] obtained *HST* photometry of asteroid 21 Lutetia in UV and visible bands. Lutetia has been classified by various authors as an M-type asteroid; in the context of this work its color index is 0.05 (making it a C type) and its $A_{\mathrm{MPC}} = 0.06$ suggest it to be a $\mathrm{C_{high}}$ type in particular, though in this work's system we formally would require a *WISE* p_{W1} measurement to classify it as such. In the following section we show that M types (a group in the Tholen taxonomic system) and our $\mathrm{C_{high}}$ types are largely the same population.

The *HST* Lutetia photometry revealed a steep drop in albedo around ∼300 nm and nearly constant albedo in the 200–300 nm region at a factor ∼0.6 times the visible (r-band-equivalent) albedo. The *HST* observations of Lutetia thus generally agree with the C-type albedo trend (Figure 5.9), the main difference being the location of the UV albedo drop-off (the bluest two ECAS bands also demonstrate this difference between M types and C types, e.g., see Figure 5.2 of [Bus et al. (2002)]).. The *Rosetta* spacecraft's flyby of Lutetia enabled FUV observations with the on-board Alice UV

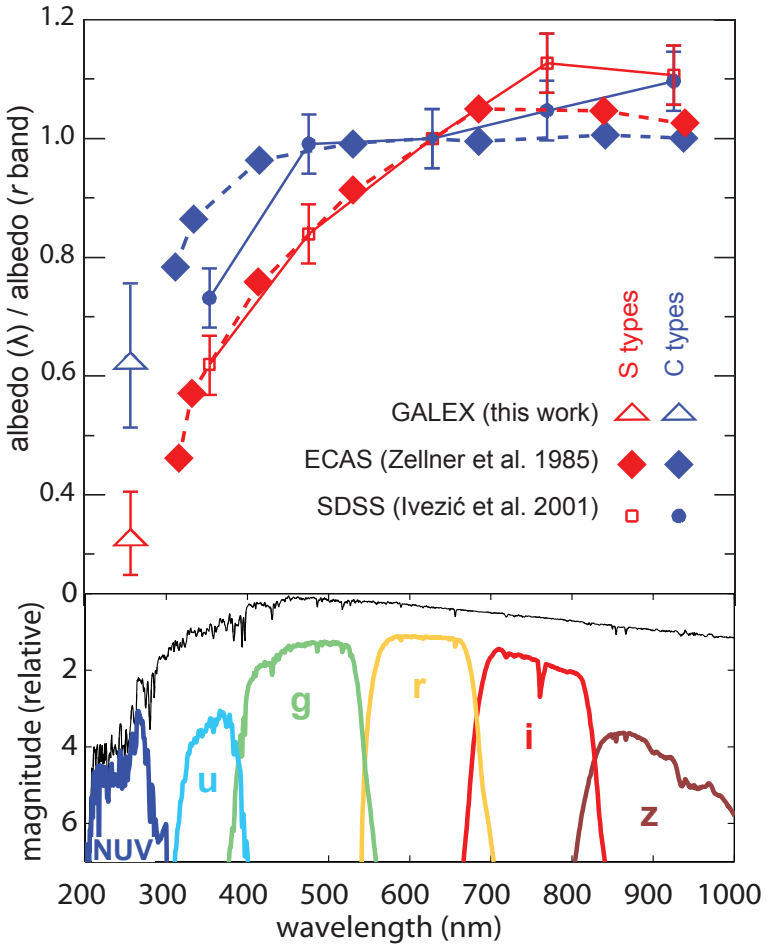


Figure 5.8: *Top*: Relative albedo (each type's r-band albedo normalized to unity) versus wavelength for C types and S types. The leftmost (NUV) points are computed from this work's data, the remaining albedos (SDSS bands) are taken directly from [Ivezić et al. (2001)]. *Bottom*: Bandpass response functions (colored lines) convolved with the solar spectrum (black line).

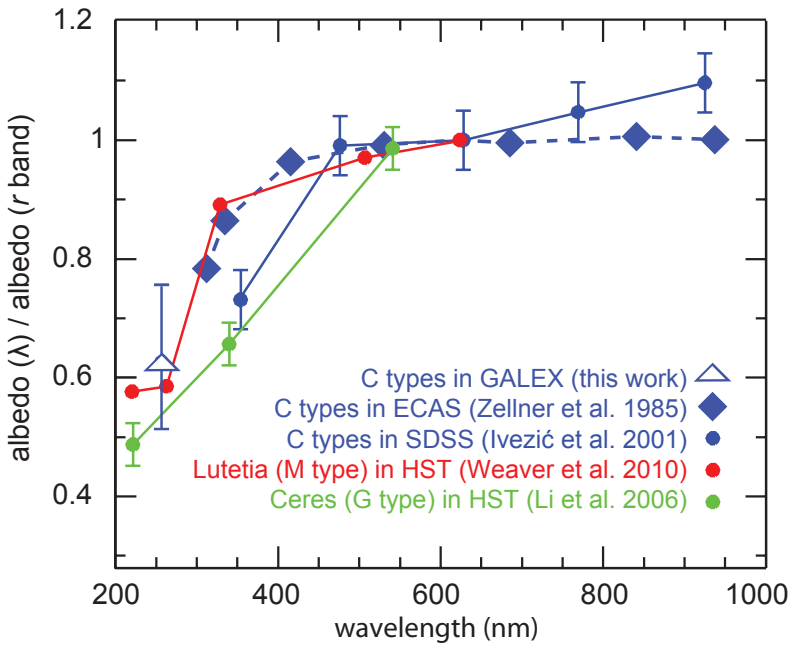


Figure 5.9: *Top*: *GALEX*, SDSS, and ECAS C-type data from Figure 5.8 compared to *HST*-derived albedos for 1 Ceres and 21 Lutetia. Note the red color here has a different meaning than it does in Figure 5.8.

imaging spectrograph [Stern et al. (2011)]; the longest wavelengths of the FUV data (∼190 nm) yield an albedo consistent with the constant value measured in the 200–300 nm range by *HST*.

5.6.2 Ceres

HST photometry of asteroid 1 Ceres has also been obtained in the UV and visible ([Parker et al. (2002)], [Li et al. (2006)]). With a color index of 0.01, Ceres is also a C type in our classification scheme, though its $A_{\rm MPC} = 0.033$ makes its placement in our $\rm C_{low}$ vs. $\rm C_{high}$ groups ambiguous (see Figure 5.6 panel F). Like Lutetia, Ceres lacks a reported p_{W1} so that we cannot formally classify it as either $\rm C_{low}$ or $\rm C_{high}$.

Ceres was observed by *GALEX* and thus is included in our MPC-data-based analysis; our measured NUV$-V_{\rm MPC} = 6.45\pm0.19$ for Ceres make it a clear outlier from the C-type NUV $- V_{\rm MPC}$ distribution (Figure 5.4 panel B). In the Tholen taxonomic system Ceres is classified as a G type; in the following section we show that other G

types exhibit similarly high NUV $- V_{\rm MPC}$ values but less anomalous NUV $- V_{\rm PTF}$.

The [Parker et al. (2002)] *HST* data show that around ∼300 nm Ceres' albedo drops to as low as ∼0.3 times the visible-band albedo—compared to the factor of ∼0.6 seen for *GALEX* C types and the Lutetia data—but that around ∼200 nm it appears to rise again to a more typical C-type UV albedo. [Roettger and Buratti (1994)] did not observe this unusually deep absorption feature near 300 nm in their *IUE* spectrum of Ceres; if real this feature could partially explain the anomalous NUV $- V_{\rm MPC}$ we observe for G types in *GALEX*. Figure 5.9 shows Ceres data in the three *HST* bands observed by [Li et al. (2006)], none of which sample the 300-nm region containing the putative absorption band, though these three bands do generally match the *GALEX* C-type data.

5.7 C-type subgroups

C types deserve further consideration for several reasons: (1) C types outnumber S types in the *GALEX* samples by a factor of several, (2) our division of C types into $\rm C_{low}$ and $\rm C_{high}$ merits interpretation in more conventional taxonomic systems, and (3) both of the *HST*-observed asteroids in the previous section are known members of C-type subgroups, the UV properties of which are worth confirming with additional group members.

Figures 10 and 11 detail the distribution of *GALEX*-observed asteroids belonging to six classes each from the Tholen and Bus/Binzel taxonomic systems ([Tholen (1989)]; [Bus & Binzel (2002)]), the latter is sometimes referred to as the SMASSII system after the survey data with which it was derived. These two classification systems were created on the basis of different visible-band color data; a comparison of their group definitions is given in Table 5.1 of [Bus et al. (2002)]. We consider only the subset of *GALEX*-observed asteroids having both $V_{\rm MPC}$ and $V_{\rm PTF}$ model magnitudes and omit subgroups containing less than three objects. In the following subsections we briefly comment on these subgroups.

One key interpretation of these data—supported also by the *HST* data in Figure

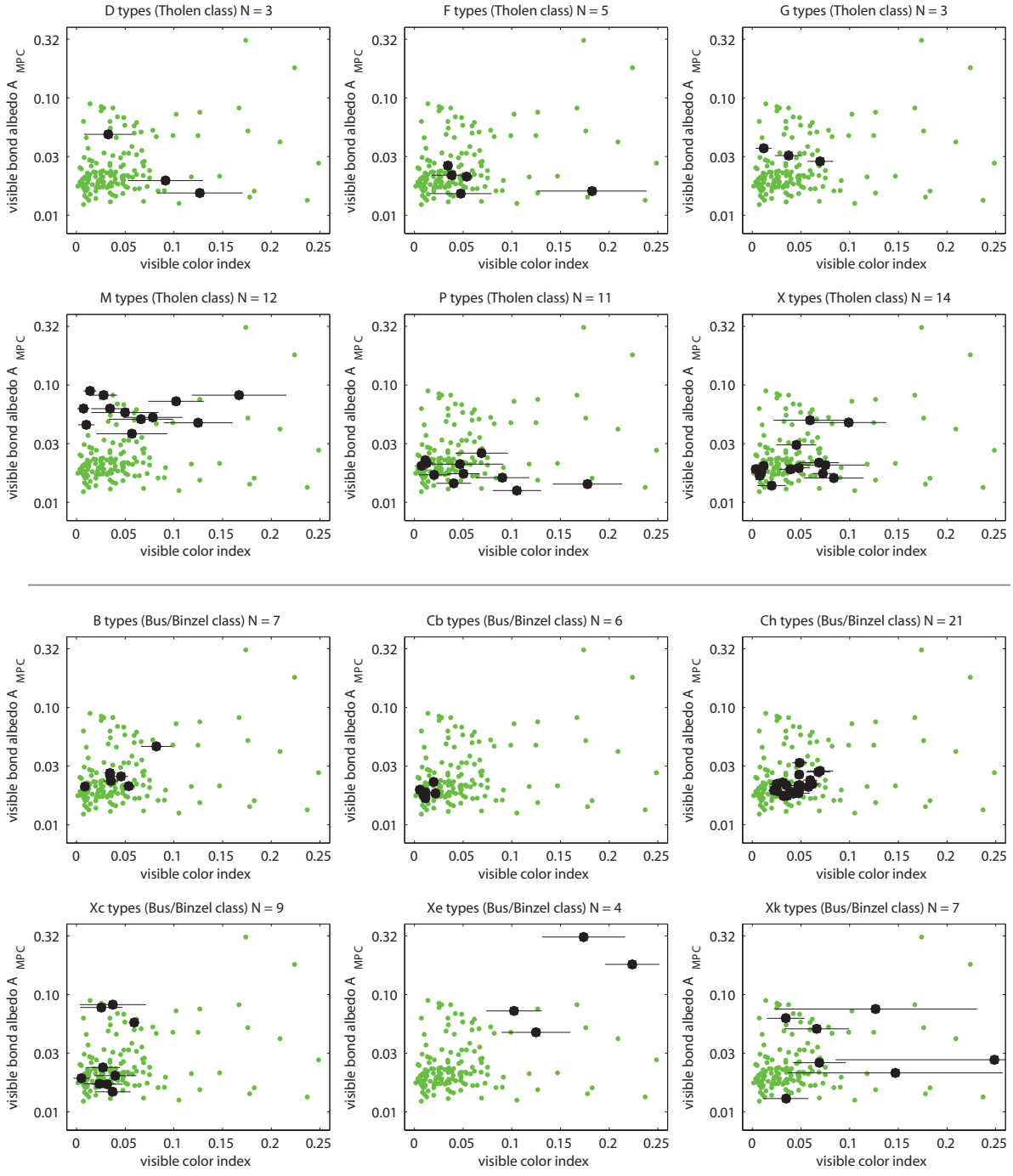


Figure 5.10: Visible color/albedo distributions of Tholen-classified [Tholen (1989)] and Bus/Binzel-classified [Bus & Binzel (2002)] C-type subgroups among this work's sample of *GALEX*-observed C-type asteroids. See text for further information.

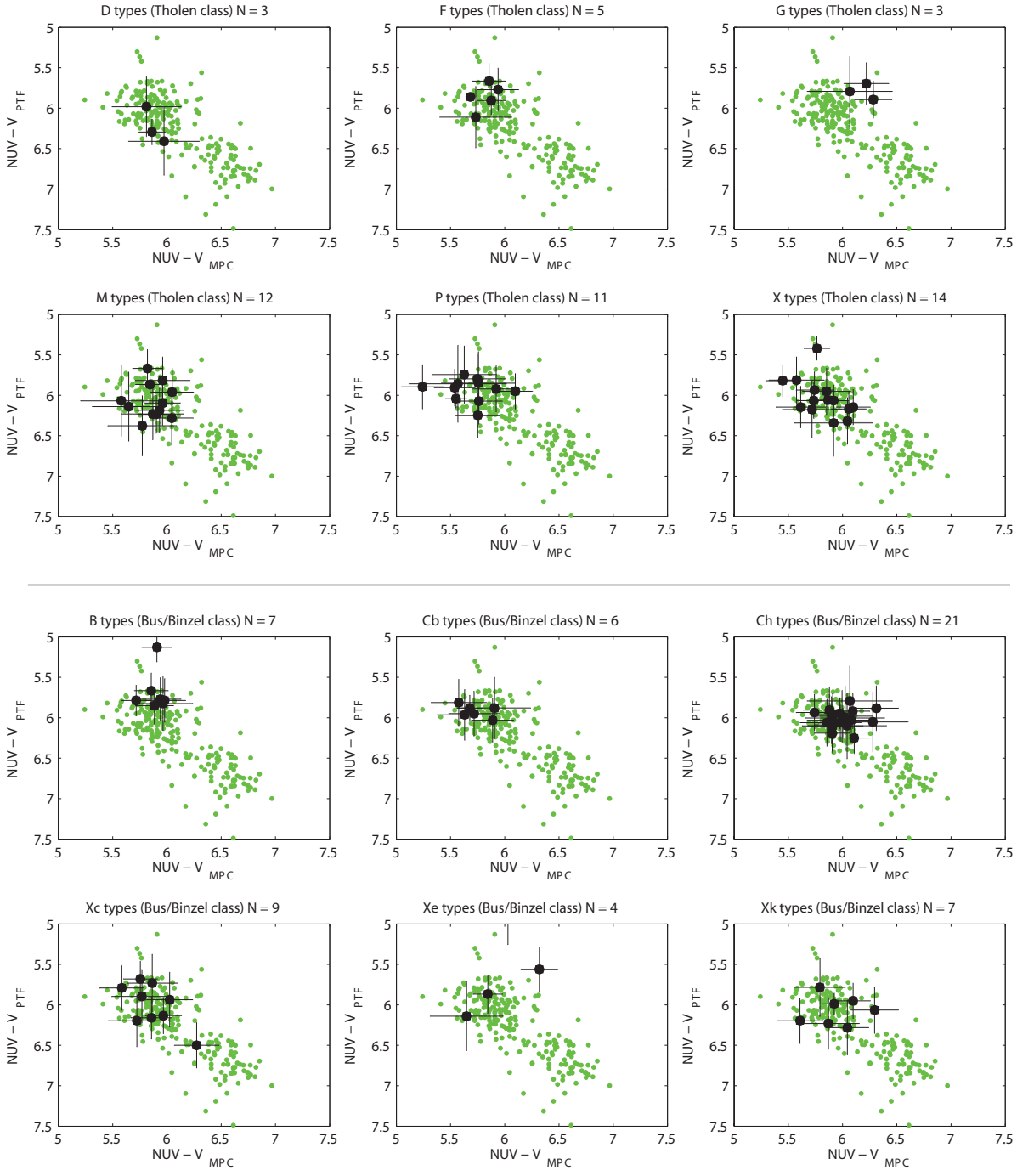


Figure 5.11: NUV $-$ V color distributions of Tholen-classified [Tholen (1989)] and Bus/Binzel-classified [Bus & Binzel (2002)] C-type subgroups among this work's sample of *GALEX*-observed C-type asteroids. See text for further information.

5.9—is that NUV-band albedo is not very useful for discriminating C-type subgroups, e.g., M types versus G types, whereas u band appears to be more diagnostic in this regard. The u-band discrepancy between these subgroups was remarked most notably by [Zellner et al. (1985)] in the ECAS data, but it was unknown at that time (indeed, up until now) whether the discrepancy in UV albedo became more or less pronounced shortward of ∼300 nm. The NUV data indicate that the discrepancy lessens in the NUV, as M types do in fact exhibit a step-down in albedo (between NUV and u bands), similar to the step down the G types exhibit within u band.

5.7.1 X complex

The Tholen system's X-type group includes asteroids with relatively flat visible color, including no substantial absorption in the blue (in contrast to, e.g., the u-band drop-off seen in G types). The subgroups within the X group include M, P and E types and are distinguishable only by albedo.

The twelve M types in our sample all have $A_{\rm MPC} > 0.03$ and $p_{W1} > 0.125$, the latter formally makes them all $\rm C_{high}$ types in this work's classification system. The M types have $\rm NUV - V_{MPC} = 5.89 \pm 0.15$ and $\rm NUV - V_{PTF} = 6.12 \pm 0.21$, neither of which significantly differ from the C-type averages given in Table 5.3. This is consistent with the above-noted observation that M-type Lutetia's $\rm NUV - V$ is similar to that of the *GALEX* C types, despite an obvious difference in u-band (Figure 5.9). Assuming *all* 29 of the $\rm C_{high}$ types in the *GALEX* sample are in fact M types, then the $\rm C_{high}$ types' slightly higher $\rm NUV - V_{PTF} = 6.14 \pm 0.33$ (compared to $\rm NUV - V_{PTF} = 6.03 \pm 0.22$ for the whole C type group) agrees well with the M types' slightly higher average.

Complementary to the M types, the eleven P types in our sample all have $A_{\rm MPC} < 0.03$ and $p_{W1} < 0.125$, the latter formally makes them all $\rm C_{low}$ types. The P types have $\rm NUV - V_{MPC} = 5.74 \pm 0.17$ and $\rm NUV - V_{PTF} = 5.91 \pm 0.13$. These values are less than both models' C-type averages as well as less than the $\rm C_{low}$ average, suggesting our $\rm C_{low}$ group includes more diverse objects than just P types (e.g., the five F types also all have $A_{\rm MPC}$ consistent with $\rm C_{low}$).

There are 14 *GALEX*-observed asteroids listed simply as X types in the Tholen system (presumably because no visible albedo was available at the time of classification); Figure 5.10 shows that these are in fact distributed across both the C_{low} and C_{high} albedo ranges.

In the Bus/Binzel system, the X complex consists of four subgroups: Xc, Xk, X and Xe, these being differentiated by their spectral slope and presence of various absorption features. In the *GALEX* sample the most numerous of these are the Xc types, which have the least red visible color and seem to include both high and low visible albedo members. Both the Xe and Xk types have higher visible color indices (with larger uncertainties in the color). As with the Tholen X types, we see no systematic trends with respect to the NUV properties of these subgroups.

5.7.2 G types

Three *GALEX*-observed asteroids are categorized as G types. Like G-type Ceres, these have intermediate $A_{\rm MPC}$ and an above-average NUV $- V_{\rm MPC} = 6.22 \pm 0.11$. In contrast, however, the G-type NUV $- V_{\rm PTF} = 5.79 \pm 0.10$ lies slightly below the C-type average. The reason for this discrepancy is that all three G types in this sample have $p_{W1} < 0.125$ and so are formally classed as C_{low} objects, as a result their assumed $A_{\rm bond} = 0.01$ in the computation of $V_{\rm PTF}$ may be too low. On the other hand, [Oszkiewicz et al. (2011)] fit $G_{12} = 0.88 \pm 0.2$ to Ceres' phase function, suggesting that the assumed $G_{12} = 0.84 \pm 0.14$ for C_{low} types (Table 5.2) is a more valid assumption for G types than the C_{high} value of $G_{12} = 0.42 \pm 0.20$. Hence the G types seem not to fit well into either of our C_{low} or C_{high} groups, and hence are not accurately modeled by our $V_{\rm PTF}$.

The three-asteroid G-type sample's higher than average NUV $- V_{\rm MPC}$ agrees with the Cere's HST data (Figure 5.9), which as discussed above could be indicative of an absorption feature at ∼300 nm unique to G types [Li et al. (2006)], the precise shape and location of which remains unresolved in the broadband photometry considered here.

Tholen's G types are represented in the Bus/Binzel system by the Cg and Cgh groups; however no asteroids in our *GALEX* sample have either of these SMASSII labels.

5.7.3 B types

Members of the Tholen B and F classes, represented also by the Bus/Binzel B and Cb classes, all are classified as C_{low} types in the *GALEX* sample based on their p_{W1}. Unlike the G types, the B types are not anomalous in NUV $- V_{\rm MPC}$, meaning the B types likely lack the G types' strong absorption at 300nm. The B types also are characterized by slightly higher $A_{\rm MPC} = 0.026$ compared to the C_{low} average $A_{\rm MPC} = 0.020$. Hence, like the G types, the B types show a lower than average NUV $- V_{\rm PTF}$ symptomatic of an underestimated $A_{\rm bond}$ and therefore too dim of a predicted $V_{\rm PTF}$.

5.8 Summary

We present NUV-band photometry of 405 asteroids observed serendipitously by *GALEX* from 2003–2012. Using a compilation of visible-band color data, we select the subset of these *GALEX*-observed asteroids belonging to the C-type or S-type classes. We then compute the visual-band magnitude (using two different models) corresponding to each *GALEX* detection in an effort to study the NUV $- V$ color. For both V models, the derived NUV $- V$ color distribution is bimodal, with S types having the redder color, just as they do within the visible band. The average C-type NUV $- V$ agrees with *HST* observations of the asteroids Lutetia and Ceres, both of which are members of the visible-color-defined C-type group. Slight differences in the measured NUV $- V$ among known taxonomic subgroups of the C types may indicate membership in either the M-type or G-type subgroups, though the 300–400 nm region (u-band) is more diagnostic of this division.

Chapter 6

Trojan Lightcurve Demographics

6.1 Introduction

[Lagrange(1772)] demonstrated that one solution to the general three-body problem is that of constant equal distances—a planet's *Trojans* are minor bodies satisfying this condition by occupying the co-orbital Lagrange points L_4 and L_5, situated 60 deg ahead of and behind the planet in its orbit. Jupiter has the largest known and most studied Trojan population, although Earth, Mars, Uranus, and Neptune also have known Trojans of varying degrees of stability ([Connors et al. (2011)]; [Bowell (1990)]; [Alexandersen et al. (2013)]; [Chiang et al. (2003)]).

According to the original version of the *Nice model* ([Morissey et al. (2005))]; [Tsiganis et al. (2005)]; [Gomes et al. (2005))]) the Jovian Trojans accumulated via chaotic capture during an instability caused by Jupiter and Saturn crossing a mutual mean-motion resonance. Later revisions to the Nice model incorporated a non-continuous resonance crossing ("jumping Jupiter"; [Brasser et al. (2009)]) and a fifth giant planet that was ejected during the instability ([Batygin et al. (2012)]; [Nesvorný & Morbidelli (2012))]). [Nesvorný et al. (2013))] showed that not only is the Trojans' capture still reproduced with these revisions, but also that the revised model predicts an asymmetry between the number of Trojans at L_4 and L_5, which the original version [Morissey et al. (2005))] did *not* predict.

The asymmetry between the two Trojan swarms is one example of a demographic trend well-suited for characterization with wide-field surveys such as the Sloan Dig-

Table 6.1: Trojan color data sets used in computing this work's C/D taxonomic metric. These data sets are visualized in Figure 6.1 panel A.

survey name	references	data description	# Trojans
Eight-Color Asteroid Survey (ECAS)	[Zellner et al. (1985)] [Zellner et al. (2009)]	Photometry in eight custom filters measured with photomultipliers at Catalina and Steward Observatories. We compute and use the principal component color index PC#1 = $0.771(b-v) - 0.637(v-w)$. Excludes objects with PC#1 error >0.3 mag.	15
Sloan Digital Sky Survey (SDSS) *griz* colors	[Ivezić et al. (2002)] [Parker et al. (2008)] [Ivezić et al. (2010)]	g,r,i, and z broadband CCD photometry acquired by SDSS from 1998–2009. Includes data in the Moving Object Catalog v4, supplemented with post-2007 detections from SDSS DR10. We use the first principal component a^* defined in the references. Excludes objects with a^* error >0.05 mag or $(i-z)$ error >0.1 mag.	269
Fornasier et al. photometric survey	[Fornasier et al. (2004)] [Fornasier et al. (2007)]	V,R, and I broadband photometry acquired at ESO (La Silla) from 2002–2005.	51
Wide-field Infrared Survey Explorer (WISE)	[Grav et al. (2011))] [Grav et al. (2012))]	$W1$-band 3.4μm (near-infrared) geometric albedo p_{W1}. Based upon using thermal ($W3$ & $W4$-band) fluxes to first estimate the diameter. Excludes objects with p_{W1} error $\geq$0.03.	141
Jewitt & Luu visible-band spectroscopy	[Jewitt & Luu (1990)]	CCD spectroscopy (0.4–0.7μm, $R \approx 300$) conducted at Kitt Peak.	30
Emery et al. near-IR spectroscopic survey	[Emery et al. (2011)]	Near-IR spectroscopy (0.7–2.5μm, $R \approx 130$) conducted at the NASA Infrared Telescope Facility (IRTF) from 2003–2007. Color defined between 0.85μm and H-band (1.6μm).	69

ital sky-survey (SDSS; [York et al. (2000)]) and the *Wide-field Infrared Survey Explorer* (WISE; [Wright et al. (2010)]). In addition to confirming the L_4/L_5 asymmetry (namely, that $\sim$50% more Trojans exist at L_4 than at L_5), these surveys confirmed the preponderance of two distinct spectral groups, recognized in earlier studies (e.g., [Jewitt & Luu (1990)]) as C types and D types. [Bus et al. (2002)] review these and other asteroid taxonomic classes, which are defined on the basis of low-resolution ($R \approx 100$) visible reflectance spectra. In this work we employ a clustering technique for distinguishing 'C type' and 'D type' Trojans, applying it to a compilation of visible-color data sets (a technique used previously in defining the C-type vs. S-type metric in [Waszczak et al. (2015)]).

The two Trojan groups may have been captured from disparate regions of our solar system's primordial planetesimal disk, where different chemical and/or physical conditions might have lead to different bulk and/or surface compositions (e.g., [Brown (2012)]). The similarity of D types to comet nuclei [Jewitt & Luu (1990)] suggested their outer solar system origin well before the discovery of the Kuiper belt and other trans-Neptunian objects—many of which also share the distinctly red colors [Ofek (2012)] of D-type Trojans. C types on the other hand are one of the two dominant groups in the main asteroid belt and likely originated in the giant-planet region (between Jupiter and Neptune). This history for main-belt C types is consistent with simulations of a hypothetical "hot Jupiter" episode in our early solar system (the *Grand Tack model*, [Walsh et al. (2011)]), which simultaneously reproduce the coarse orbital and compositional structure of the main asteroid belt.

No variants of the Nice model or Grand Tack simulations have yet been used (or attempted) to reproduce the Trojans' dual-composition. Ideally such a model would not only trace the source regions of the Trojans' two compositional groups, but should predict any orbital structure, number-size distributions, spin properties, and/or surface roughness characterizing either group given their dynamical histories. Observational programs are thus tasked with elucidating any such key demographic trends. These might include (or suffer contamination from) population trends acquired post-emplacement, e.g., due to collisions or the effects of solar radiation over

several billions of years.

[Grav et al. (2012))] for instance showed (with the WISE sample) that the ratio of the two Trojan types varies as a function of diameter—a conclusion independently verified by [Wong et al. (2014)]. [Sonnett et al. (2015)] used WISE lightcurves to investigate the binarity fraction of Trojans, another clue to the origin and history of small-body populations (e.g., [Goldreich et al. (2002))]) that should be further explored within the Trojan subgroups. Orbital trends may also exist within the groups: for example, [Emery et al. (2011)] noticed a possible correlation between near-IR color and inclination within the C-type group. A taxonomic asymmetry may exist between the L_4 and L_5 swarms [Roig et al. (2008)]; however this result depends on the controversial definition of Trojan dynamical families [Brož & Rozehnal (2011)].

The present work examines the rotational (spin) and solar-phase-angle-dependent photometric properties of Trojans as inferred from lightcurve observations. We append derived parameters from 107 Trojan lightcurves appearing in the larger (predominantly main-belt) lightcurve set of [Waszczak et al. (2015)] to 80 additional Trojan lightcurves referenced from the literature.

6.2 Color-based taxonomic metric

I adopt the multi-data-set *fuzzy c-means* (FCM) clustering method described in the appendix of [Waszczak et al. (2015)]. In that work, the method was applied to all *main-belt* asteroids appearing in at least one of seven colorimetry data sets—here, we apply it to all *Trojans* appearing in at least one of six data sets. Whereas for main-belt asteroids the two clusters considered were the C types and S types, in this work the clustering analysis distinguishes C types from D types.

The six data sets are described in Table 5.1. As seen in Figure 6.1 panel A, the first three data sets are two dimensional, while the last three are one dimensional. Note that in the case of [Waszczak et al. (2015)], all data sets used were two dimensional; the FCM clustering technique is equally applicable to 1D data as it is to higher dimensional data. Regardless of the dimensionality, the number of clusters to be

identified is all data sets is input as two, given prior knowledge of the two color groups' existence. Correlations between the output cluster-membership scores (for each Trojan, a probability between 0 = C-type and 1 = D-type) is seen in the off-diagonal plots in Figure 6.1.

Compiling these various data sets not only serves to average the predicted class membership of Trojans occurring in multiple surveys, it also yields a unified sample having greater completeness across a wider range of diameters. Targeted surveys such as the Eight-Color Asteroid Survey or the IRTF survey of [Emery et al. (2011)] tend to include only the brightest objects, while wide-field surveys like SDSS and WISE contribute serendipitous data on many smaller objects.

The colorimetric or spectroscopic observations comprising these six surveys all are indicators of the spectral slope ('redness') of the Trojans included therein. For this reason, some authors have referred to the D and C-type Trojans more descriptively as the 'redder' and 'less-red' groups ([Emery et al. (2011)], [Wong et al. (2014)]). Some authors also identify P types as members of one of the groups or as their own, intermediate-redness group. [Szabó et al. (2007)] resolved the two compositional groups in SDSS using the color index $t^* = 0.93(r-i) + 0.34(i-z) - 0.25$. The width and location of the r, i and z bands make t^* a proxy for the spectral reflectance slope between 0.7 μm and 1 μm. Here for the SDSS data instead of t^* we use the more commonly-cited color index $a^* = 0.89(g-r) + 0.45(r-i) - 0.57$, which was derived to maximize separation of main-belt C and S types and is a measure of the slope between $\sim$0.4μm and 0.8μm. For the WISE sample, Grav et al. ([Grav et al. (2011))], [Grav et al. (2012))]) reported each Trojan's albedo in the near-infrared W1 band (3.4 μm), denoted p_{W1}. Technically this single-band reflectance is, by itself, not a color measurement. However, given that the visible band albedos of the two types are very similar ($p_V \approx 0.06$ for both types), the p_{W1} is indeed a proxy for the visible-to-NIR reflectance slope.

Of the 6,277 Trojans known as of this analysis (size distribution in Figure 6.1 panel A; diameters are computed assuming $p_V = 0.06$ for all Trojans), 423 are classifiable with this work's taxonomic metric based on their inclusion in one or more of the six

data sets. We define D types as those with color indices greater than 2/3 (271 objects) and C types as those with color indices less than 1/3 (128 objects). There are 24 Trojans with color indices between 1/3 and 2/3 which are ambiguously classified—either because they are equally distant from both clusters in one or more surveys, or because they occupy conflicting clusters in two or more surveys and thus average to near 1/2.

6.3 Rotation periods and amplitudes

[Waszczak et al. (2015)] reviews the basics of using lightcurve observations to extract information about the shape and spin periods of asteroids, as applied in particular to the Palomar Transient Factory (PTF) set of ∼9,000 lightcurves. For a description of the period-fitting algorithm to which the PTF lightcurves were subjected, the algorithm's performance, table of output lightcurve parameters, and an analysis of the main-belt asteroid lightcurve statistics, we refer the reader to that work. In the PTF lightcurve set there are 117 Trojan lightcurves (of 100 unique Trojans) that have 'reliable' period fits, where 'reliable' is defined in detail in that work. Of those 100, there are 38 Trojans that also have a taxonomic classification according to this work's cluster-based scheme. Figure 6.1 panel B shows the number-size distribution of Trojans with periods in PTF and color information.

Plots of only a small selection of PTF Trojan lightcurves appear in Figures 6.2, 6.7 and 6.8 of this chapter. Plots of the remaining ∼100 PTF lightcurves may be generated using the data tables available in [Waszczak et al. (2015)].

The PTF Trojan lightcurve sample is dominated by smaller ($D < 40$ km) Trojans. To increase the sample size overall and the completeness at the large-diameter end, we supplement these with data from the Light Curve Data Base (LCDB, [Warner et al. (2009)]), an online repository for lightcurve parameters contributed by a large community of observers over many decades. We consider only those Trojans with quality code 3 (highest reliability score) periods. This increases the lightcurve sample to 164 Trojans (86 with $D > 40$ km), of which 93 (66 with $D > 40$ km) have a

taxonomic color index available. Some works contributing lightcurves to the Trojan LCDB data include [Mann et al. (2007)], [Mottola et al. (2011))], [Melita et al. (2010)], French et al. ([French et al. (2011)], [French et al. (2012)], [French et al. (2015)]), and [Warner & Stevens (2011)]. Each contributing observer/program describes their own observational program, data reduction and time-series analysis in their respective works; all contributed lightcurve data include a period estimate and a lower limit estimate of the amplitude.

The middle column of Figure 6.2 shows the distribution of the 164 Trojans in diameter, period, and amplitude space. The running geometric mean and one-sigma percentiles are also overlaid as green and red lines, respectively, to aid visualization of the change with diameter. For all data points in these plots, the diameter is computed from the absolute visual magnitude (H) and an assumed albedo $p_V = 0.06$. The right-hand column of plots show the subset of these data that have taxonomic labels (color indices), and the separate geometric means for the two groups. The left column of plots in Figure 6.2 compares these same data to the distribution of *main-belt* C types, shown in green in the plot backgrounds. These main-belt C types are also the union of PTF lightcurve data and LCBD data (the latter contributes most of the $D > 100$ km data for main-belt C types). Furthermore, the PTF main-belt C types include asteroids classified as such on the basis of their photometric index [Waszczak et al. (2015)], as opposed (or in addition) to their color index. Though some D types have been found in the main-belt [DeMeo et al. (2013)], these are far too rare to consider in the same population-level analysis we undertake here.

The bottom middle and bottom right panels of Figure 6.2 include contours describing the theoretical maximal amplitudes of self-gravitating cohesionless rotators, of various bulk densities, associated with spinning at the break-up limit. [Waszczak et al. (2015)] showed for example that main-belt asteroids exhibit a sharp upper limit of $\sim$2 g/cm^3 for S types and a slightly lower upper limit of 1-2 g/cm^3 for C types. Again, these are only upper limits under the assumption that we are seeing cohesionless bodies that happen to be rotating at their maximum possible rate. All of the Trojan data points fall to the left of the 0.5 g/cm^3 contour, meaning

that—under these same assumptions—the Trojans could have bulk densities as low as 0.5 g/cm^3. The only specific Trojans having density estimates available by other methods are both binary systems. The first such system is 617 Patroclus, measured by [Marchis et al. (2006)] as 0.8 g/cm^3 and by [Muller et al. (2010))] as 1.08 ± 0.33 g/cm^3. The second is 624 Hektor, measured by [Marchis et al. (2014)] as 1.0 ± 0.3 g/cm^3.

Figure 6.3 presents the same data from Figure 6.2 albeit omitting the individual points and only showing the running geocentric mean and running 16th–84th percentile (1σ) contours. Apart from differences with respect to smaller-scale structure, for $D > 40$ km the Trojans qualitatively match the main-belt C types very well in terms of both period and amplitude. We verify this formally in Figure 6.4 by comparing the one-dimensional distributions of spin rates and amplitudes (for $D > 40$ km objects only) with a two-sided Kolmogorov-Smirnov (KS) test [Massey (1951))]. The Trojan C types and main-belt C types are not statistically different distributions to more than 2σ, meaning we cannot reject the null hypothesis that they are indeed the same distribution. The same conclusion hold for the comparison of large Trojan D types with large Trojan C types, and for comparing both Trojan types together against the main-belt.

In making the one-dimensional histograms in Figure 6.4, we plot the best-fitted Maxwellian distribution for visual reference; however we have not performed the usual per-datum normalization by the running geometric mean in these plots. Even lacking that normalization, the $D > 40$ km objects do fit a Maxwellian frequency distribution reasonably well, certainly better than do the $D < 40$ km objects (Figure 6.5), a property shared by the main-belt population as described by [Pravec & Harris (2000)] and [Pravec et al. (2002)]. [Waszczak et al. (2015)] noted the main-belt asteroids' qualitative fit to a Maxwellian as well, but showed that the observed $D > 40$ km distribution actually differs from its best-fit Maxwellian at the 10σ significance level (via a KS test). Here, the 86-object $D > 40$ km Trojan sample differs at about the 9σ level from a Maxwellian.

A Maxwellian distribution in spin frequencies has often been cited as the evolved

state for a population of collisionally-evolved particles (e.g., [Salo (1987)]), this being most easily understood as a vector combination of Gaussian-distributed angular momenta distributions in each of the three spatial dimensions. Recent work [Steinberg & Sari (2015)] suggests that a Lévy distribution may be a more accurate model, and that the apparent lack of the data's fit to a Lévy distribution may indicate a significant residual primordial (or at least non-collisional) spin component. [Mottola et al. (2014))] consider the possibility that the large Trojans' deviation from a Maxwellian may be evidence for episodes of nonisotropic cometary outgassing during a Trojan's lifetime, the angular momentum change of which dominates over the mostly isotropic nature of the collisions, which otherwise would create the Maxwellian distribution but instead introduces more spread in the distribution (both fast and slow rotators).

At 10–40 km sizes, the Trojan and main-belt populations differ in the following ways:

1. unlike the main-belt asteroids, small Trojans, i.e., $10 < (D/\mathrm{km}) < 40$ objects, do *not* show an increased fraction of fast rotators piling up at a ∼2-hour spin barrier. Rather, the 84th percentile spin rate of smaller Trojans is the same as for larger Trojans (∼6 hours),

2. small Trojans show a greater proportion of slow rotators (periods of several days to tens of days) as compared to similar-sized main belt objects,

3. small Trojans have an apparently higher mean amplitude than do the main-belt C types, though the 84th percentile of the two populations' amplitudes is remarkably identical over the observed size range.

Regarding point (1) above, it is difficult to infer whether the observed 5–6 hour lower limit on period is itself a spin barrier analogous the main-belt's ∼2 hour barrier, or whether it is a manifestation of some other mechanism or event, perhaps primordial and possibly shared with the main-belt C types (and still presently among the $D > 40$

km populations). If the former (a spin barrier), then as mentioned earlier it could imply Trojan bulk densities of as low as $\sim$0.5 g/cm^3.

[French et al. (2015)] discuss the possible role of the Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect in the differing spin distributions for small and large Trojans. YORP is the gradual change in asteroids' angular momentum vectors (direction and magnitude) due to asymmetric torques created by photon pressure on their irregularly-shaped surfaces. The forces involved may only become significant over several billions of years. For YORP, the change in spin rate is inversely proportional to both the square of the orbital semi-major axis and the square (or higher power) of the object's linear size. An asteroid twice as far from the sun experiences 1/4 the spin rate change, as does an object twice the size at the same orbital radius.

Again as noted by [French et al. (2015)], if Trojans and main-belt C types are the same bulk density, then the largest Trojans affected by YORP at 5 AU should be half the size of the smallest susceptible main belt objects. Taking the latter to be $\sim$40 km, then only Trojans smaller than 20 km will have YORP-influenced spin states. Though this work's sample does not show an excess of faster rotators near or below $D = 20$ km, note the sample's completeness at this size is only of order 1%. On the other hand, if $\sim$5 hours is indeed the fastest possible spin period for Trojans, then the addition of YORP forces at 20 km cannot produce a population of faster rotators (they would break apart), it could however contribute to an excess of slow ($\sim$10-day period) spin rates, which is indeed observed. Slow rotators may also be, in some cases, binary systems that have drained their angular momentum through tidal locking. Finally, the extent to which the cometary outgassing scenario of [Mottola et al. (2014))] might be applied to $D < 40$ km Trojans is unclear, however given that cometary activity has been seen in $D < 10$ km main-belt objects (the main-belt comets, e.g., [Jewitt (2012)]), the capacity for similar sized Trojans to retain sub-surface ice reservoirs is not unreasonable.

6.4 Binarity

Lightcurves of binary asteroid systems are typically characterized by amplitudes of $\Delta m > 0.9$ mag, with deep 'V'-shaped or cusp-like minima. These minima correspond to the mutual event (eclipse or occultation) occurring once per rotation period (if the two members are contacting) or orbital period (if detached). As mentioned above, binary systems can provide opportunities to estimate the asteroids' bulk densities given precise (i.e., adaptive-optics) astrometric observations (e.g., [Marchis et al. (2006)]). Their population frequency may offer clues about small-body formation and dynamical evolution, as has been explored for example in the context of KBOs (e.g., [Goldreich et al. (2002))]).

[Sonnett et al. (2015)] further motivates the search for Trojan binaries, and uses the *WISE*-observed sample to identify 34 candidate binary Trojans. Six of these happen to have lightcurves in the PTF sample (Figure 6.7). One of these (asteroid 13323) has two lightcurves in PTF: one from the 2010 opposition and the other from 2014. In both of these oppositions, its fitted lightcurve full-amplitude is less than 0.5 mag, contrary to the 0.92 ± 0.08 mag observed by *WISE*. Not only is he 2014 PTF lightcurve is lacking data at the minima, it could also represent a different (non-eclipsing) viewing geometry. The 2010 PTF lightcurve of 13323 *is* however from the same opposition as the *WISE* lightcurve. On the other hand, the *WISE* lightcurve clearly shows two minima of different depths; it is certainly possible that the deeper of these two is undersampled in PTF, resulting in the underestimated PTF amplitude.

[Sonnett et al. (2015)] did not comment on the taxonomic types of the binary candidates they identified. Nine out of the 34 candidates they identified have a color index in the taxonomic metric we use in this work; according to these colors seven of the *WISE* binary candidates are D types and two are C types.

Because it parameterizes period-folded rotation curves using only the four coefficients of a second-order Fourier series, the PTF lightcurve-fitting algorithm [Waszczak et al. (2015)] was neither designed nor reliable for the successful identification or modeling of binary-asteroid lightcurves. Given sufficiently densely and

evenly sampled lightcurves in PTF, the fitting procedure sometimes outputs a 'valid' lightcurve solution for binaries that may fit the maxima well but give significant residuals with respect to the minimum-flux observations. Assuming there are indeed positive detections at the minima, these data points will, in successive iterations, receive inflated uncertainty values (errorbars) in an effort to minimize the fit's chi-squared. If the errorbars grow too large, the minima data points can often be discarded, and the lightcurve fitting may be aborted before a desirable set of fitted parameters is obtained. However, in some instances the fraction of data occurring at minima is sufficiently low, and the rest of the lightcurve adequately constrained such that a solution is produced and the lightcurve is added to the sample.

We visually identified two likely binaries among the PTF lightcurves (Figure 6.8): asteroids 5244 and 9430. The latter has three lightcurves in PTF, both R and g band from 2012 and R band from 2013. Their color indices indicate that asteroid 5244 is a C-type Trojan and 9340 is a D type. The lightcurves in Figure 6.8 do exhibit the cusp-like minima described above, in contrast to the *WISE* binary candidates in Figure 6.7.

6.5 Photometric phase functions

As with rotational properties, [Waszczak et al. (2015)] provide a detailed overview of asteroids' photometric variation as a function of solar-phase-angle (i.e., phase functions). Figure 6.1 of [Waszczak et al. (2015)] includes example C-type (main-belt) and D-type (Trojan) phase function data from the literature; we also produced nearly 4,000 new fitted phase functions (mostly main-belt asteroids) from the PTF data. In PTF we fit three phase function models to lightcurves, two of which we consider here for the Trojans: the G_{12} model of [Muinonen et al. (2010))], and a simpler linear model β (units of mag/deg).

In general, asteroid phase-function lightcurves are much more time consuming to produce from targeted observations as compared to rotation lightcurves, given the much longer observational baseline required. PTF and other wide-field surveys

provide an unprecedentedly efficient means of collecting asteroid phase-curve data. Unfortunately, this also means that, unlike the case of rotational parameters, the literature does not contain a very large pre-existing catalog of phase parameter measurements (compared to the size of the rotational database).

[French (1987)] was the first to publish a Trojan-asteroid phase curve; she found that (1173) Anchises exhibits a very linear phase curve lacking the commonly-seen opposition effect. Anchises' phase function data therefore were poorly fit by the [Bowell et al. (1989)] G-parameter phase function model. Despite being classified as a C-type Trojan according to this work's color index, Anchises' observed phase function differs from the typically-fitted $G \approx 0.08$ of main-belt asteroids. We note however that the [French (1987)] data were limited to phase angles of less than 2 deg (a very limited sampling compared to main belt asteroids, which are typically observed out to 20–30 deg).

[Schaefer et al. (2010)] observed nine Trojans and fit both a multi-parameter (Hapke) and a single-parameter (linear) phase function mode. They claim the linear model yielded comparable fits to the Hapke model (i.e., similar χ^2), with most fitted slopes in the range $0.04 \leq \beta \leq 0.06$, albeit with stated uncertainties of 20-40% and one object with a negative ($S < 0$) fitted slope. A key shortcoming in the analysis of [Schaefer et al. (2010)] was their lack of a fit for the rotational modulation, which, as we explain in [Waszczak et al. (2015)], causes underestimated uncertainties in the fitted parameters.

Another work [Shevchenko et al. (2012)] observed three Trojans and computed higher-precision linear phase function fits. They accurately fit the rotational component with high-cadence sampling at each epoch (or, equivalently, at each phase angle), and computed phase-slope (β) values of 0.045, 0.040 and 0.044 (all $\pm$0.001 mag/deg, or 2-3% uncertainties) for 588 Achilles, 884 Priamus and 1143 Odysseus, respectively. All three are classified as D-type Trojans in the color-index-based taxonomic classification scheme.

In Figure 6.9 we plot the fitted G_{12} and β parameters for 23 taxonomically-classified Trojans that have sufficient phase-angle coverage (data occupying three

or more phase-angle bins of width 3 deg) and either a reliable fitted period *or* a fitted amplitude of less than 0.1 mag. Neither the G_{12} nor β distributions of the C types are statistically distinct from the D types, nor are the Trojan C types distinct from the much larger main-belt C types in their phase function parameters.

More measurements of Trojan phase curves are required to better discern the extent to which the C and D types (and main-belt and Trojan asteroids) differ in their phase function parameters. For main-belt S and C types, [Waszczak et al. (2015)] showed that the two populations clearly differ in their phase function, though this does not necessarily mean, for instance, that all (color-index-defined) C types throughout the solar system should have the same phase function. As noted in [Waszczak et al. (2015)], the phase function depends on many physical surface properties (grain size, compactness, topographic roughness, and more) which could conceivably vary independently of the bulk (or at least surface) mineralogical/chemical composition of the asteroids to which we attribute the observed colors. Differing collisional rates, ambient dust densities, solar radiation levels and other orbital environmental factors may lead to differing photometric properties that might ultimately be seen in the phase-function parameter (but perhaps not color) distributions.

A related point is that the phase function of an asteroid is commonly recognized of as being intricately related to the object's albedo. As we have been assuming the two Trojan types (and Trojans and main-belt C types) have a common geometric albedo of $p_V = 0.06$, the hypothesis that they could exhibit differences in phase function may seem inconsistent. However, [Waszczak et al. (2015)] details how the geometric albedo, while formally *defined* in terms of the phase function, is not necessarily an unambiguous proxy for the phase function; they demonstrate this with main-belt asteroid data in particular.

6.6 Summary

We present a sample of rotational (period and amplitude) parameters for 164 Trojan lightcurves, and use it to search for differences between the two Trojan taxonomic types and between Trojans to main-belt C types. This sample is a combination of new lightcurves from the PTF survey (mostly $D < 40$ km objects) and existing lightcurves from the literature (primarily $D > 40$ km and larger objects, considering only those with quality code 3 periods). We also consider binarity and the distribution of phase-function parameters. Any potential discrepancies in these populations' lightcurve properties would need to be taken into account by dynamical and other models of the origin and evolution of these small bodies.

In order to label Trojans as either C or D type, we employ a color-based taxonomic scheme based upon a clustering analysis of six different data sets, in a manner similar to what was done for main-belt asteroids in [Waszczak et al. (2015)].

The large Trojans ($D > 40$ km) show no statistically significant differences in the spin period or spin amplitude distributions, both in comparing C and D types and Trojans and main-belt asteroids. The samples in this size range are ∼40% complete for C and D types, ∼50% complete for Trojans irrespective of color type, and ∼30% complete for main-belt C types. Because these results are based on KS-test comparisons, we can only report a non-detection of any difference in their distributions (as opposed to confidence that they are in fact identical distributions, which we cannot claim).

For smaller Trojans, the spin and amplitude distributions do differ at the 3–4σ level. While the sample sizes in this regime are far smaller in terms of population completeness, some key differences stand out, including the lack of Trojans rotating fast than ∼5 hours. Whether this represents a 'spin barrier' analogous to the main-belt's ∼2-hour limit is not clear; if true it could imply the Trojans have bulk densities as low as ∼0.5 gm/cm^3, comparable to the low densities determined by astrometric observations of known Trojan binaries.

The effects of YORP would probably not be seen on Trojans larger than 10–

20 km in size, assuming a simple inverse-square relationship with semi-major axis and comparing to the main-belt population. This smaller influence of YORP relative to main-belt asteroids was proposed by [French et al. (2015)], while a cometary outgassing scenario was suggested by [Mottola et al. (2014))].

Although the PTF lightcurve sample as currently analyzed is not ideal for the identification or study of binary systems, we do present PTF lightcurves of six Trojans that were identified by [Sonnett et al. (2015)] in *WISE*, and we identify two PTF-observed Trojans whose lightcurves are strongly suggestive of binary systems. Finally, the phase-function parameter distributions of Trojans do not show any statistically significant differences; the need for further exploration of this issue in particular is very clear, given that the phase function can conceivably contain information unrelated to the bulk composition (such as surface roughness or regolith compactness).

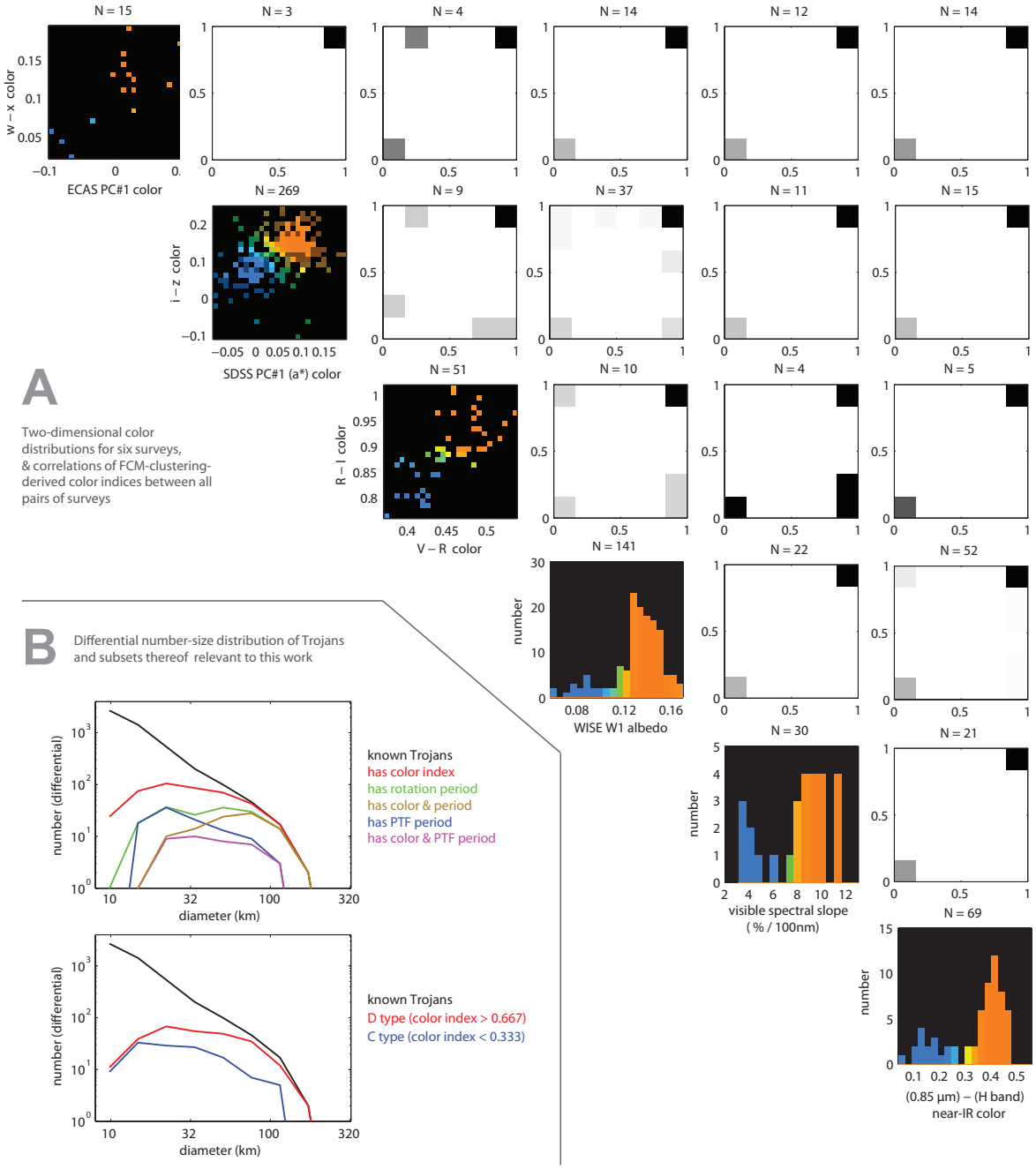


Figure 6.1: *Panel A*: Color distributions for six surveys (diagonal, with blue = C type and orange = D type), and correlations (off diagonal) of FCM-clustering-derived classifications between all pairs of surveys (all pairs of data sets share at least three Trojans). Compare to Figure 5.26 of [Waszczak et al. (2015)]. *Panel B*: Differential number size distribution of Trojans, with subsets relevant to this work. Diameters assume a visible geometric albedo of 0.06 for all Trojans.

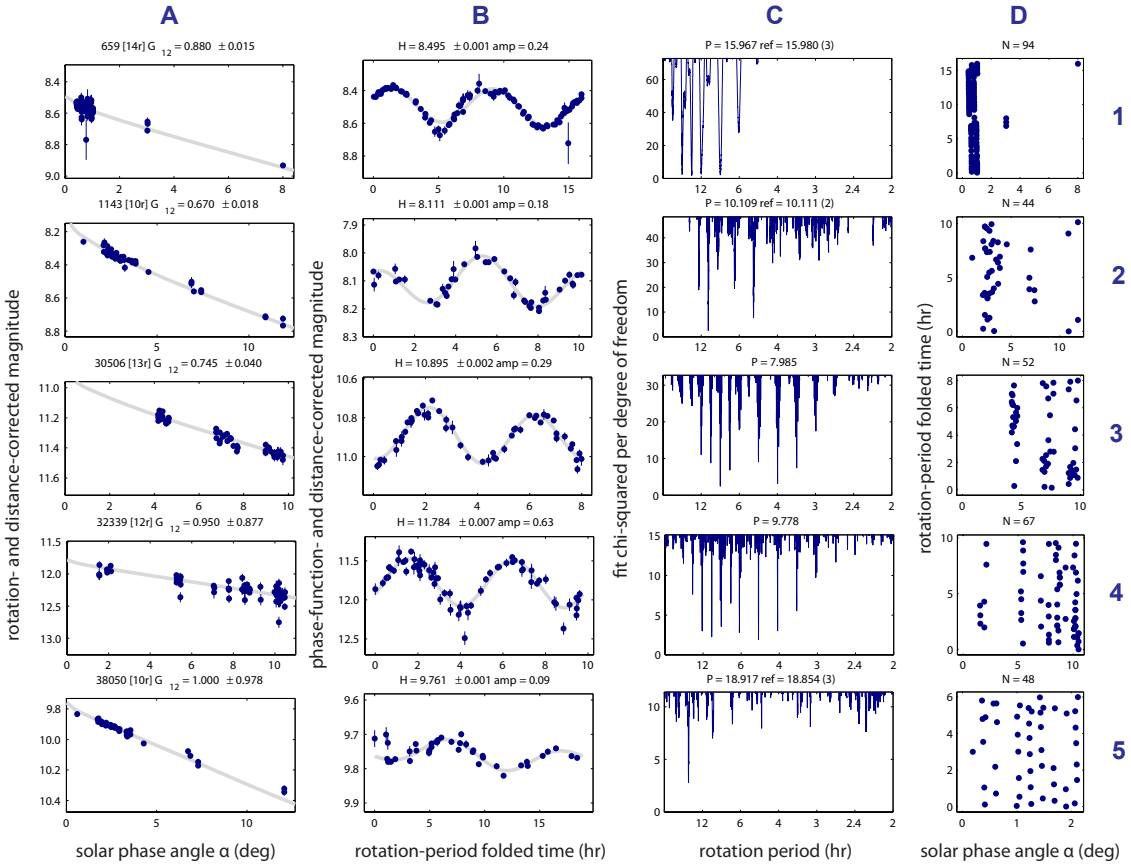


Figure 6.2: A small selection of example Trojan PTF lightcurves. Each row corresponds to a different asteroid. Column A shows the phase curve (corrected for rotation); Column B shows the rotation curve (corrected for phase-function); Column C shows the periodogram; Column D shows the distribution of the observations in rotational phase vs. solar phase angle. Above each plot is additional information depending on the column: (A) the asteroid number, followed by (in square brackets) the opposition year (most are 2013) and filter (in all cases 'r') followed by the fitted G_{12} parameter; (B) the fitted absolute magnitude and amplitude; (C) the fitted period (in hours); (D) the number of data points included (and shown) in the fit.

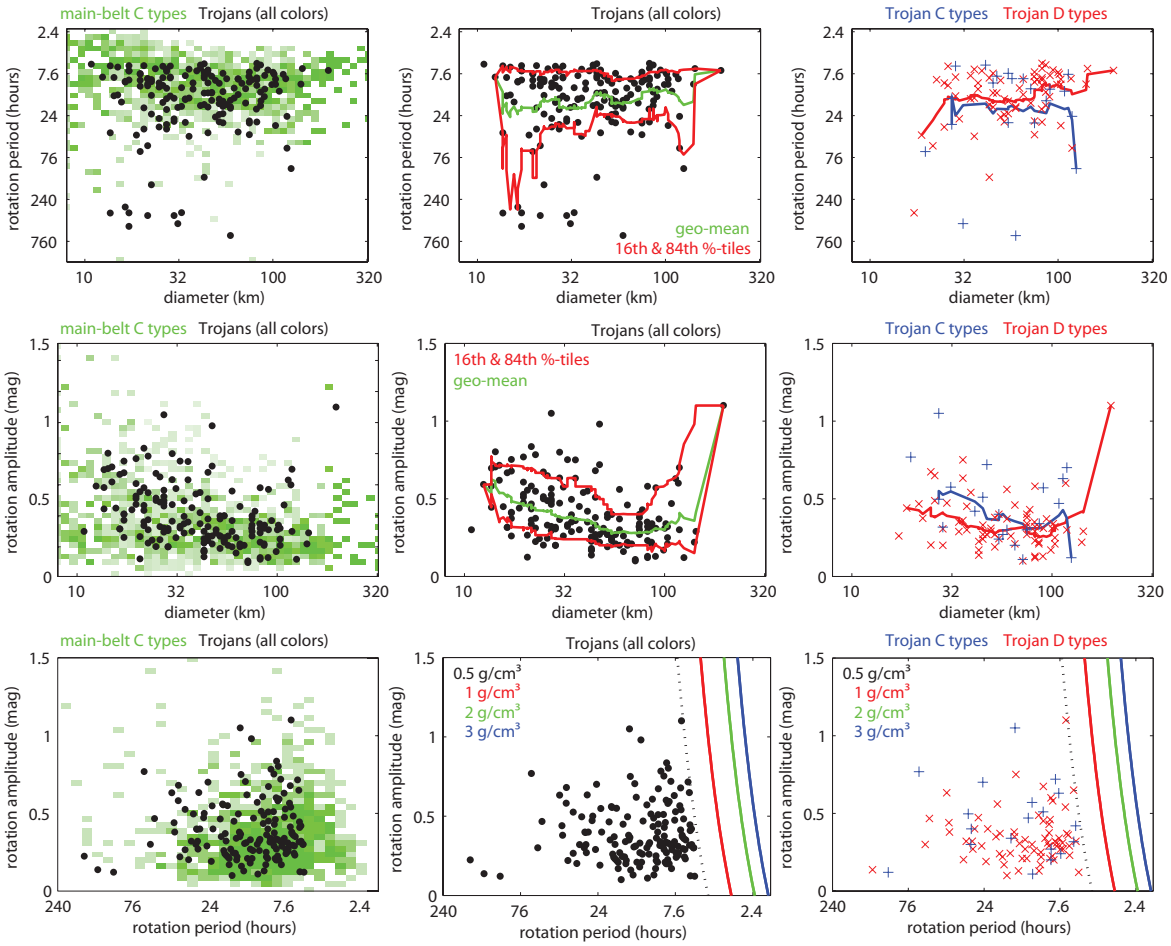


Figure 6.3: Distributions of rotation frequency and amplitude for this work's sample of Trojans (combined PTF and literature-referenced data). Diameters assume albedo $p_V = 0.06$; C-type and D-type labels are assigned based on the taxonomic metric (color index) described in Section 5.2. Lines in the bottom row's plots (period-vs.-amplitude) represent the maximum deformation states for fluid bodies of various bulk densities.

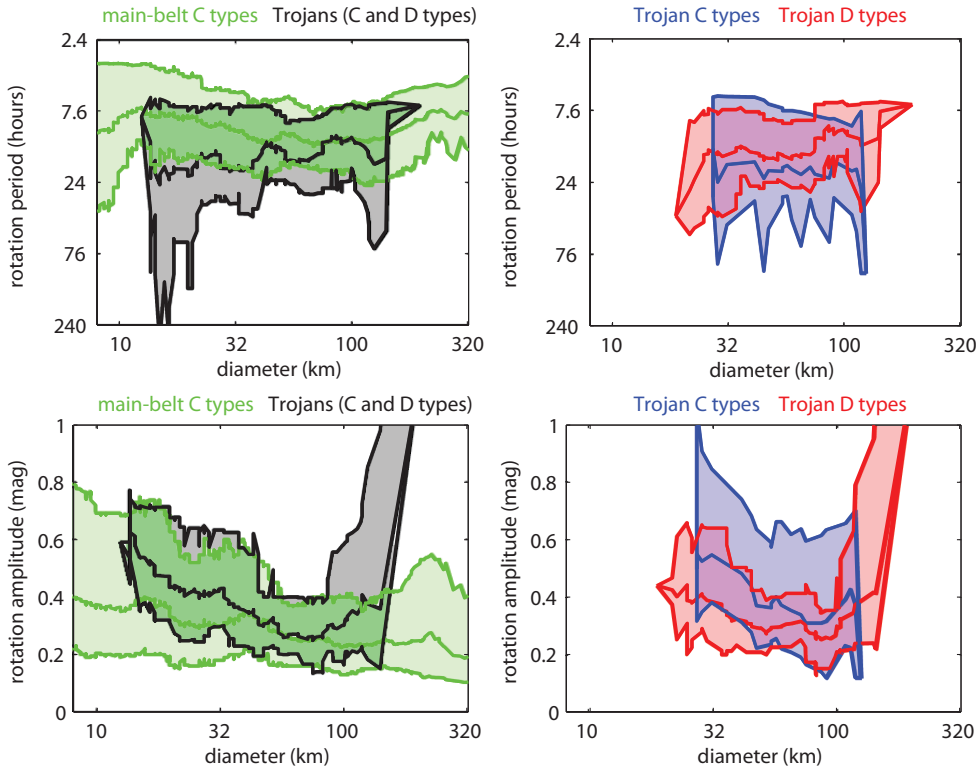


Figure 6.4: Same data as Figure 6.2, except showing only the geocentric mean and 16th–84th percentile (1σ) running statistics as functions of diameter.

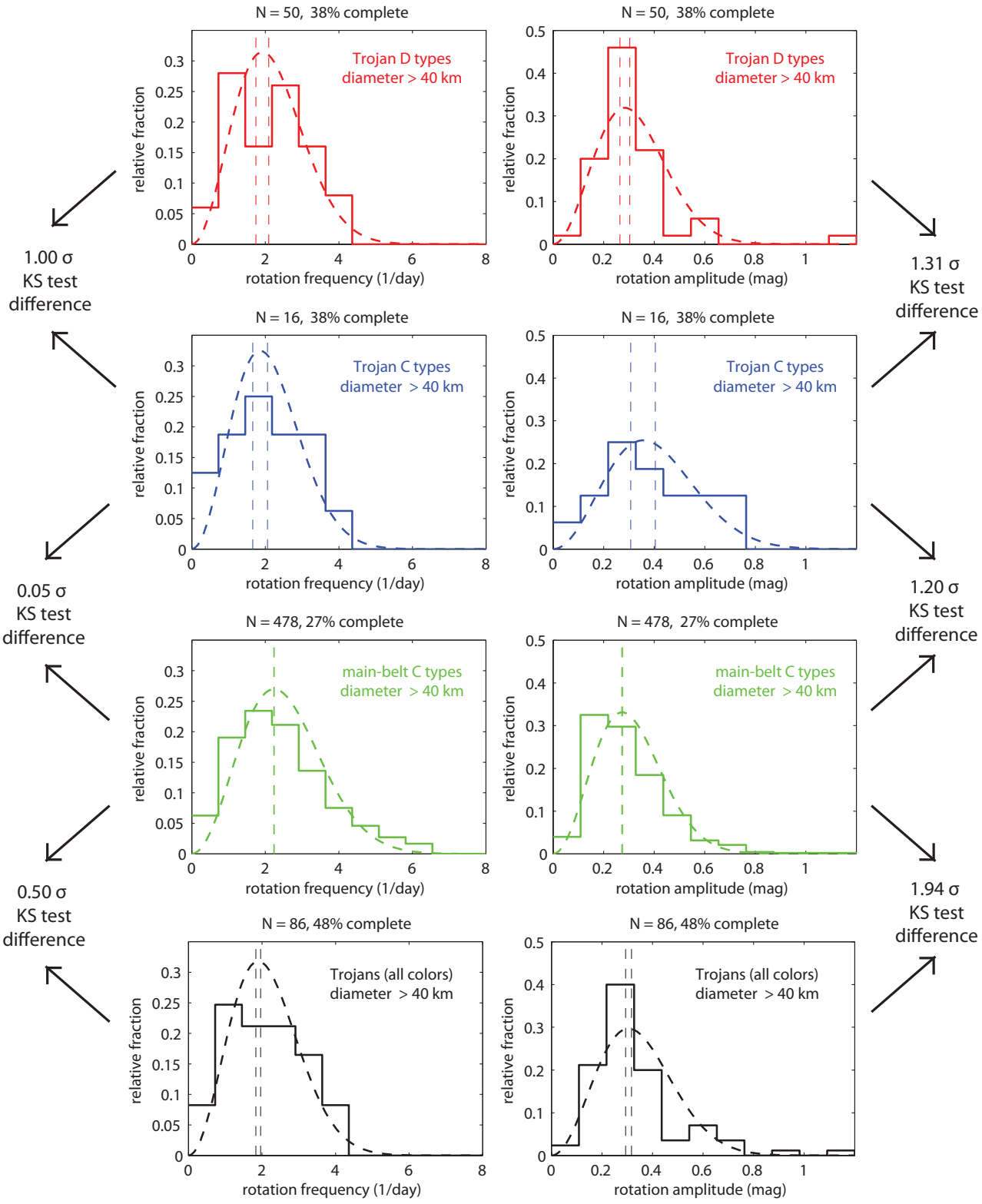


Figure 6.5: Spin frequency and amplitude distributions (not normalized per datum, as is often done) of large Trojans and main-belt C types. None of the six KS tests noted in this figure represented a statistically significant discrepancy between a pair of distributions.

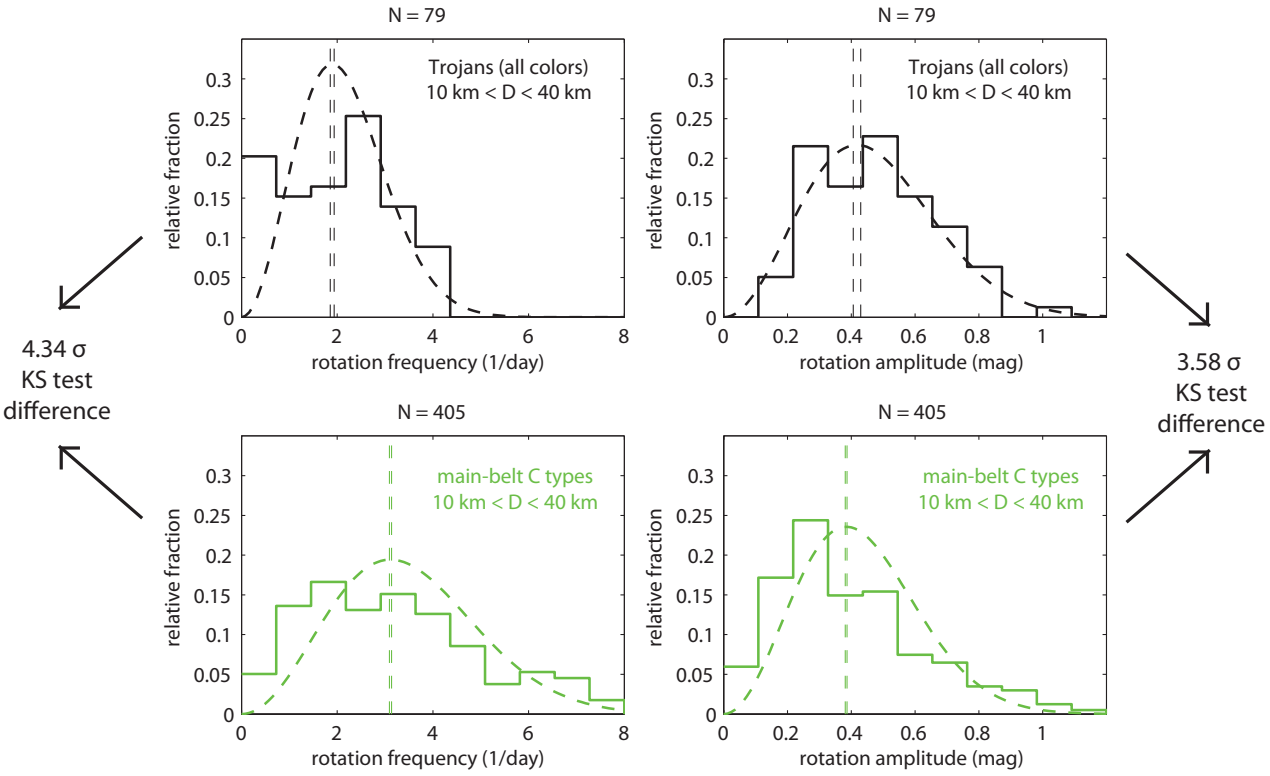


Figure 6.6: Distributions of rotation frequency and amplitude for small i.e., $10 < (D/\mathrm{km}) < 40$ size Trojans and main-belt C types. These KS tests do indicate that their observed distributions differ at least at the 3σ significance level. Trojans would probably begin to show effects of YORP in their spin statistics below ∼20 km assuming that main belt C types do around ∼40 km. There could therefore be many sub-20-km Trojans spinning with faster than ∼5 hour periods, simply beyond the diameter limit of this sample. Or, if ∼5 hours is itself the spin limit for Trojans, their observed excess in small slow rotators could also be explained by YORP at these sizes.

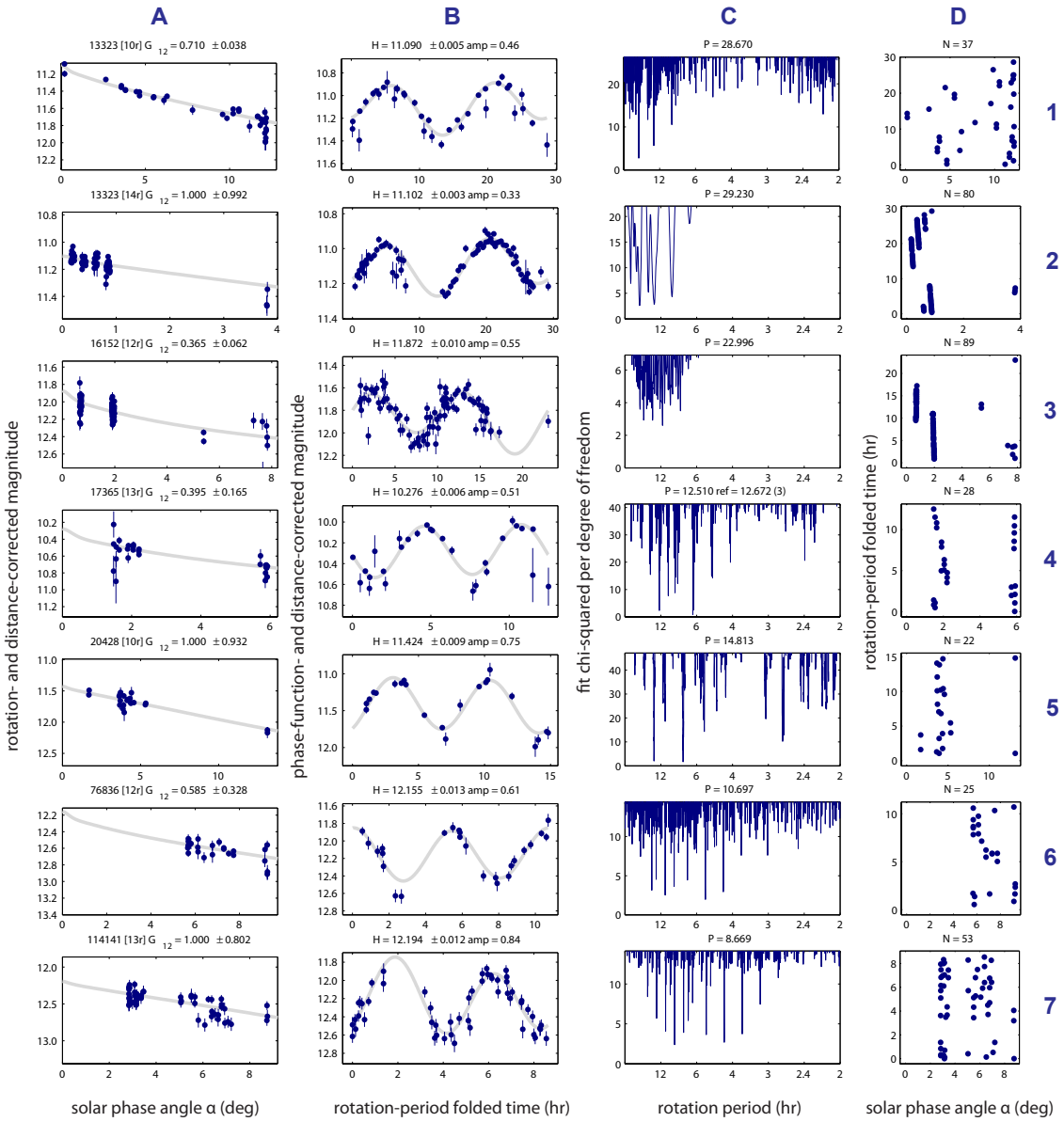


Figure 6.7: PTF lightcurves of six Trojans which were identified as candidate binaries in *WISE* (there were 34 such candidates found by [Sonnett et al. (2015)]). See Figure 6.2 for a description of the plot-labeling format used here. Note the first asteroid (13323) has two lightcurves in PTF, from different oppositions.

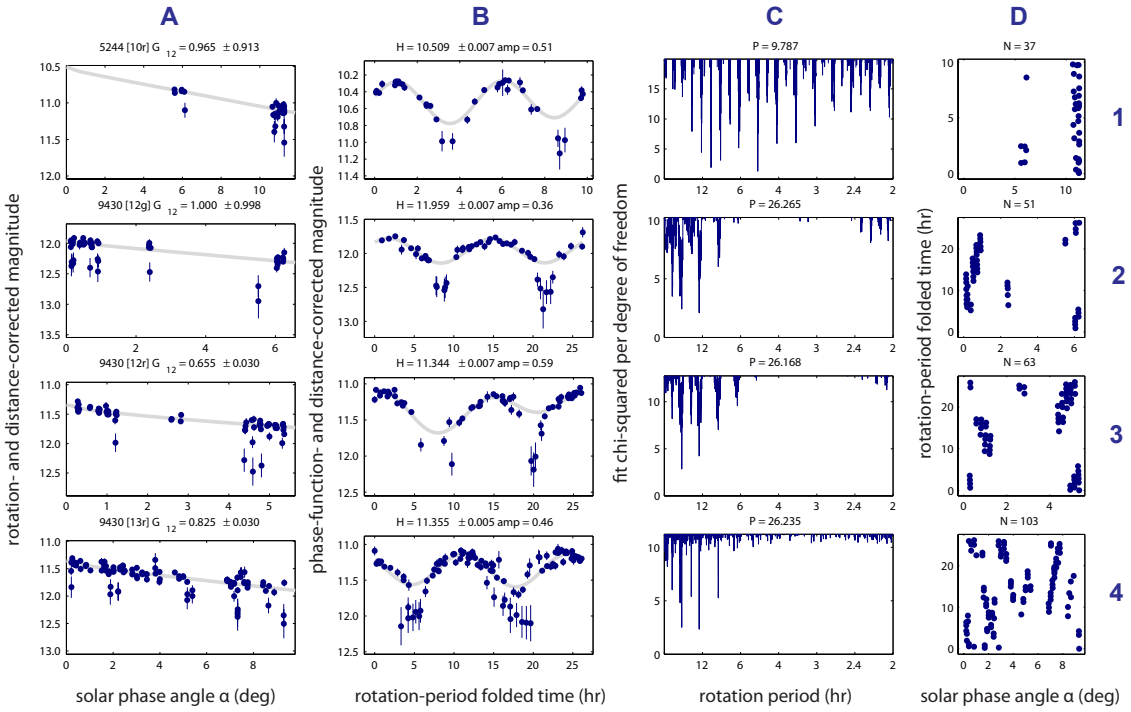


Figure 6.8: PTF lightcurves of two Trojans identified visually, in the course of this work, as likely binary systems. The second object, asteroid 9430, has three different lightcurves: two from the 2012 opposition (R and g band) and one from the 2013 opposition (R band). Note that none of the *WISE* candidates (Figure 6.8) unambiguously show the deep cusp-like minima of the kind seen in the above Trojans.

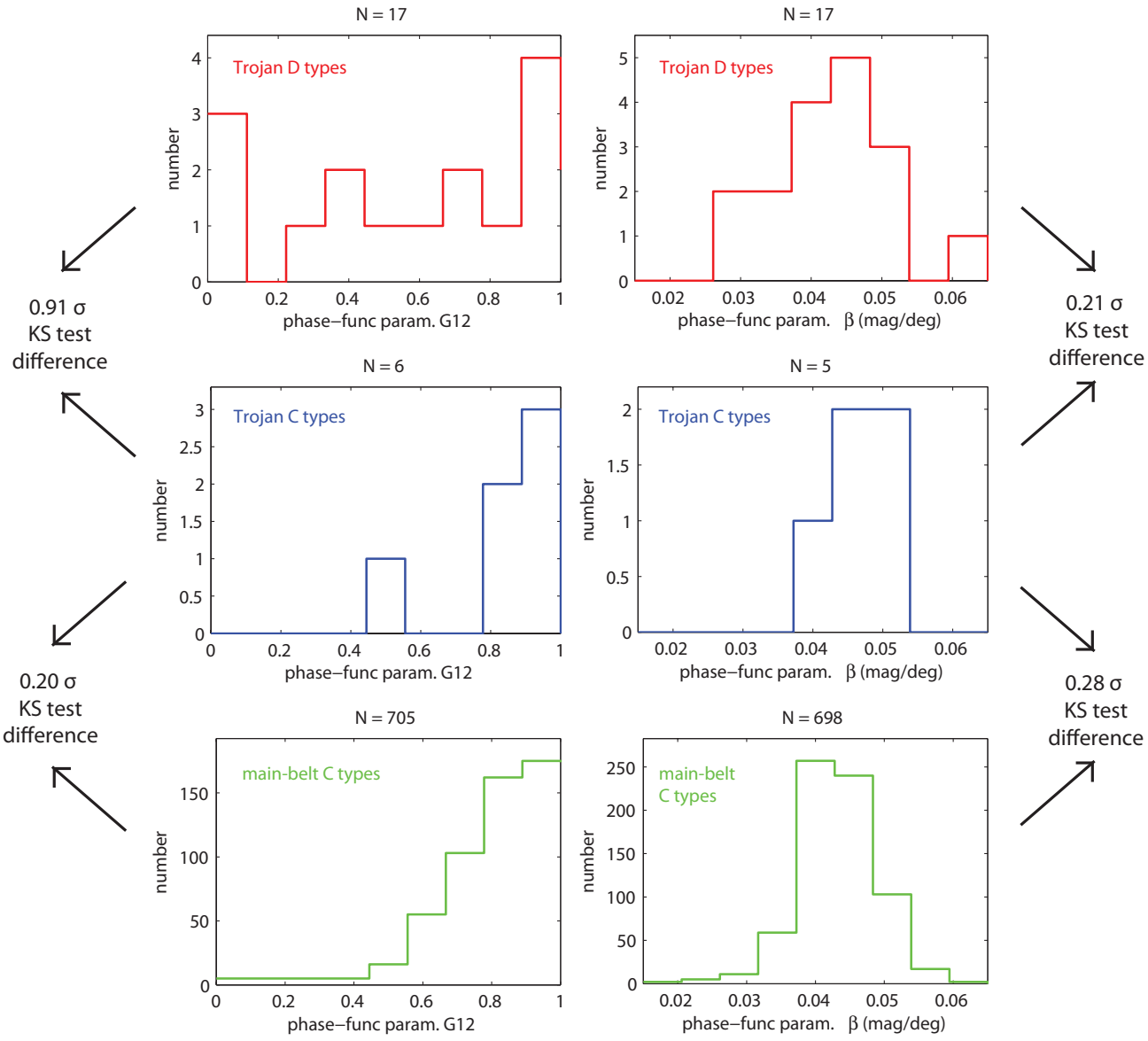


Figure 6.9: Phase function parameter distributions of PTF-observed Trojans and main-belt C-type asteroids. Only the 23 Trojans included here had sufficient phase-angle coverage and reliable period (or low amplitude) giving them a 'reliable' phase-parameter estimate.

Bibliography

[Abazajian et al. (2009)] Abazajian, K. N. et al., 2009. The seventh data release of the Sloan Digital sky-survey. http://dx.doi.org/10.1088/0067-0049/182/2/543 Astrophys. J. Suppl. 182, 543–558.

[Ahrens and Harris (1992)] Ahrens, T.J., Harris, A.W., 1992. Deflection and fragmentation of near-Earth asteroids. http://dx.doi.org/10.1038/360429a0 Nature 360, 429–433.

[Albarède (2009)] Albarède, F., 2009. Volatile accretion history of the terrestrial planets and dynamic implications. http://dx.doi.org/10.1038/nature08477 Nature 461, 1227–1233.

[Alexandersen et al. (2013)] Alexandersen, M., Gladman, B., Greenstreet, S., Kavelaars, J.J., Petit, J.-M., 2013. The first known Uranian Trojan and the frequency of temporary giant-planet co-orbitals. http://dx.doi.org/10.1126/science.1238072 Science 341, 994–997.

[Alvarez et al. (1980)] Alvarez, L.W., Alvarez, W., Asaro, F., Michel, H.V., 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. http://dx.doi.org/10.1126/science.208.4448.1095 Science 208, 1095–1108.

[Angelopoulos (2008)] Angelopoulos, V., 2008. The THEMIS mission. http://dx.doi.org/10.1007/s11214-008-9336-1 Space Sci. Rev. 141, 5–34.

[Batygin et al. (2012)] Batygin, K., Brown, M.E., Betts, H., 2012. Instability-driven dynamical evolution of a primordially five-planet outer solar system. http://dx.doi.org/10.1088/2041-8205/744/1/L3 Astrophys. J. 744, L3.

[Bauer et al. (2012)] Bauer, J.M. et al., 2012. WISE/NEOWISE observations of active bodies in the main belt. http://dx.doi.org/10.1088/0004-637X/747/1/49 Astrophys. J. 747, 49.

[Belskaya & Shevchenko (2000)] Belskaya, I.N., Shevchenko, V.G., 2000. Opposition effect of asteroids. http://dx.doi.org/10.1006/icar.2000.6410 Icarus 147, 94–105.

[Bertin and Arnouts (1996)] Bertin, E., Arnouts, S., 1996. SExtractor: Software for source extraction. http://dx.doi.org/10.1051/aas:1996164 Astron. Astrophys. Suppl. Ser. 117, 393–404.

[Bezdec (1981)] Bezdec, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

[Birtwhistle et al. (2010)] Birtwhistle, P., Ryan, W. H., Sato, H., Beshore, E. C., Kadota, K., 2010. Comet P/2010 A2 (LINEAR). http://www.cbat.eps.harvard.edu/iauc/09100/09105.html IAU Circ. 9105.

[Blomme et al. (2011)] Blomme, J., Sarro, L.M., O'Donovan, F.T., et al., 2011. Improved methodology for the automated classification of periodic variable stars. http://dx.doi.org/10.1111/j.1365-2966.2011.19466.x Mon. Not. R. Astron. Soc. 418, 96–106.

[Bloom et al. (2012)] Bloom, J., Richards, J.W., Nugent, P.E., Quimby, R.M., Kasliwal, M.M., Starr, D.L., Poznanski, D., Ofek., E.O., Cenko, S.B., Butler, N.R., Kulkarni, S.R., Gal-Yam, A., Law, N., 2012. Automating discovery and classification of transients and variable stars in the synoptic survey era. http://dx.doi.org/10.1086/668468 Publ. Astron. Soc. Pac. 124, 1175–1196.

[Bodewitts et al. (2011)] Bodewitts, D., Kelley, M. S., Li, J.Y., Landsman, W. B., Besse, S., A'Hearn, M. F., 2011. Collisional excavation of asteroid 596 Scheila. http://dx.doi.org/10.1088/2041-8205/733/1/L3 Astrophys. J. 733L, L3.

[Bolin et al. (2014)] Bolin, B., Jedicke, R., Granvik, M., Brown, P., Howell, E., Nolan, M.C., Jenniskens, P., Chyba, M., Patterson, G., Wainscoat, R.,

2014. Detecting Earths temporarily-captured natural satellites - Minimoons. http://dx.doi.org/10.1016/j.icarus.2014.05.026 Icarus 241, 280–297.

[Bottke et al. (2006)] Bottke, W.F. et al., 2013. The Yarkovsky and YORP effects: Implications for asteroid dynamics. http://dx.doi.org/10.1146/annurev.earth.34.031405.125154 Annu. Rev. Earth Planet Sci. 34, 157-191.

[Bowell (1990)] Bowell, E., 1990. 1990 MB. http://www.cbat.eps.harvard.edu/iauc/05000/05067.html IAU Circ. 5067.

[Bowell et al. (1978)] Bowell, E., Chapman, C.R., Gradie, J.C., Morrison, D., Zellner, B., 1978. Taxonomy of asteroids. http://dx.doi.org/10.1016/0019-1035(78)90085-4 Icarus 35, 313–335.

[Bowell et al. (1989)] Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., Harris, A., 1989. Application of photometric models to asteroids. In: http://adsabs.harvard.edu/abs/1989aste.conf..524B Binzel, R.P., Gehrels, T., Shapley Matthews, M.. (Eds.), Asteroids II. Univ. of Arizona Press, Tucson, pp. 524–556.

[Brasser et al. (2009)] Brasser, R., Morbidelli, A., Gomes, R., Tsiganis, K., Levison, H.F., 2009. Constructing the secular architecture of the solar system II: the terrestrial planets. http://dx.doi.org/10.1051/0004-6361/200912878 Astron. and Astrophys. 507, 1053–1065.

[Breiman (2001)] Breiman, L., 2001. Random Forests. http://dx.doi.org/10.1023/A:1010933404324 Machine Learning 45, 5–32.

[Breiman & Cutler (2004)] Breiman, L., Cutler, A., 2004. Random Forests^{TM}. http://oz.berkeley.edu/ breiman/RandomForests/ http://oz.berkeley.edu/ breiman/RandomForests/.

[Breiman et al. (1984)] Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.I., 1984. Classification and Regression Trees (Belmont, Calif.: Wadsworth).

[Brett et al. (2004)] Brett, D.R., West, R.G., Wheatley, P.J., 2004. The automated classification of astronomical light curves using Kohonen self-organizing maps. http://dx.doi.org/10.1111/j.1365-2966.2004.08093.x Mon. Not. R. Astron. Soc. 353, 369–376.

[Brož & Rozehnal (2011)] Brož, M., Rozehnal, J., 2011. Eurybates — the only asteroid family among Trojans? http://dx.doi.org/10.1111/j.1365-2966.2011.18420.x Mon. Not. R. Astron. Soc. 414, 565–574.

[Brophy et al. (2012)] Brophy, J., Culick, F., Friedman, L., Allen, C., Baughman, D., Bellerose, J., Betts, B., Brown, M.E., Busch, M., Casani, J., Coradini, M., Dankanich, J., Dimotakis, P., Elvis, M., garrick-Bethel, I., Gershman, B., Jones, T., landau, D., Lewicki, C., Lewis, J., Llanos, P., Lupisella, M., Mazanek, D., Mehrotra, P., Nuth, J., Parkin, K., Schweickart, R., Singh, G., Strange, N., Tantardini, M., Wilcox, B., Williams, C., Williams, W., Yeomans, D., 2012. Asteroid retrieval feasibility study. http://www.kiss.caltech.edu/study/asteroid/asteroid_final_report.pdf Keck Institute for Space Studies, California Institute of Technology, Pasadena, CA.

[Brown (2012)] Brown, M.E., 2012. The compositions of Kuiper belt objects. http://dx.doi.org/10.1146/annurev-earth-042711-105352 Annu. Rev. Earth Planet. Sci. 40, 467–494.

[Brown et al. (2002)] Brown, P.G., Spalding, R.E., ReVelle, D.O., Tagliaferri, E., Worden, S.P., 2002. The flux of small near-Earth objects colliding with the Earth. http://dx.doi.org/10.1038/nature01238 Nature 420, 294.

[Brown et al. (2013)] Brown, P. G., Assink, J. D., Astiz, L., Blaauw, R., Boslough, M. B., Borovička, J., Brachet, N., Brown, D., Campbell-Brown, M., Ceranna, L., Cooke, W., de Groot-Hedlin, C., Drob, D. P., Edwards, W., Evers, L. G., Garces, M., Gill, J., Hedlin, M., Kingery, A., Laske, G., Le Pichon, A., Mialle, P., Moser, D. E., Saffer, A., Silber, E., Smets, P., Spalding, R. E., Spurn, P., Tagliaferri, E., Uren, D., Weryk, R. J., Whitaker, R., Krzeminski, Z., 2013.

A 500-kiloton airburst over Chelyabinsk and an enhanced hazard from small impactors. http://dx.doi.org/10.1038/nature12741 Nature 503, 238–241.

[Burns & Safronov (1973)] Burns, J.A., Safronov, V.S., 1973. Asteroid nutation angles. http://adsabs.harvard.edu/abs/1973MNRAS.165..403B Mon. Not. R. Astron. Soc. 165, 403-411.

[Bus & Binzel (2002)] Bus, S.J., Binzel, R.P., 2002. Phase II of the Small Main-Belt Asteroid Spectroscopic Survey: The Observations. In: http://dx.doi.org/10.1006/icar.2002.6857 Icarus 158, 106–145.

[Bus & Binzel (2003)] Bus, S.J., Binzel, R.P., 2003. Small Main-belt Asteroid Spectroscopic Survey, Phase II. EAR-A-I0028-4-SBN0001/SMASSII-V1.0 http://sbn.psi.edu/pds/resource/smass2.html NASA Planetary Data System.

[Bus et al. (2002)] Bus, S.J., Vilas, F., Barucci, M., 2002. Visible-wavelength spectroscopy of asteroids. In: http://adsabs.harvard.edu/abs/2002aste.conf..169B Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 169–182.

[Bus et al. (2002)] Bus, S.J., Vilas, F., Barucci, M., 2002. Visible-wavelength spectroscopy of asteroids. In: http://adsabs.harvard.edu/abs/2002aste.conf..169B Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 169–182.

[Carbognani (2010)] Carbognani, A., 2010. Statistical analysis of C and S main-belt asteroids. http://dx.doi.org/10.1016/j.icarus.2009.08.001 Icarus 205, 497–504.

[Carry (2012)] Carry, B., 2012. Density of asteroids. http://dx.doi.org/10.1016/j.pss.2012.03.009 Planet. Space Sci. 73, 98–118.

[Cellino et al. (2009)] Cellino, A., Hestroffer, D., Tanga, P., Mottola, S., Dell'Oro, A., 2009. Genetic inversion of sparse disk-integrated photometric data of asteroids: Application to Hipparcos data. http://dx.doi.org/10.1051/0004-6361/200912134 Astron. Astrophys. 506, 935–954.

[Chambers (1999)] Chambers, J., 1999. A hybrid symplectic integrator that permits close encounters between massive bodies. http://dx.doi.org/10.1046/j.1365-8711.1999.02379.x Mon. Not. R. Astron. Soc. 304, 793–799.

[Chandrasekhar (1969)] Chandrasekhar, S. 1969. Ellipsoidal figures of equilibrium. New Haven: Yale University Press.

[Chang et al. (2014a)] Chang, C., Ip, W., Lin, H., Cheng, Y., Ngeow, C., Yang, T, Waszczak, A., Kulkarni, S.R., Levitan, D., Sesar, B., Laher, R., Surace, J., Prince, T.A., 2014. 313 New Asteroid Rotation Periods from Palomar Transient Factory Observations. http://dx.doi.org/10.1088/0004-637X/788/1/17 Astrophys. J. 788, 17.

[Chang et al. (2014b)] Chang, C., Waszczak, A., Lin, H., Ip, W., Prince, T.A., Kulkarni, S.R., Laher, R., Surace, J., 2014. A new large super-fast rotator: (335433) 2005 UW163. http://dx.doi.org/10.1088/2041-8205/791/2/L35 Astrophys. J. lett. 791, L35.

[Chapman & Gaffey (1979)] Chapman, C.R., Gaffey, M.,1979. Reflectance spectra for 277 asteroids. In: http://adsabs.harvard.edu/abs/1979aste.book..655C Asteroids (A80-24551 08-91) Tucson, Ariz., University of Arizona Press, 655-687.

[Chapman et al. (1993)] Chapman, C.R., Gaffey, M., and McFadden, L., 1993. 24-color Asteroid Survey. EAR-A-DBP-3-RDR-24COLOR-V2.1. http://sbn.psi.edu/pds/resource/24color.html NASA Planetary Data System.

[Chiang et al. (2003)] Chiang, E.I. et al., 2003. Resonance occupation in the Kuiper belt: Case examples of the 5:2 and Trojan resonances. http://dx.doi.org/10.1086/375207 Astron. J. 126, 430–443.

[Chiu (1994)] Chiu, S., 1994. Fuzzy model identification based on cluster estimation. http://dx.doi.org/10.3233/IFS-1994-2306 Journal of Intelligent and Fuzzy Systems 2, 267–278.

[Cikota et al. (2014)] Cikota, S., Ortiz, J.L., Cikota, A., Morales, N., Tancredi, G., 2014. A photometric search for active main belt asteroids. http://dx.doi.org/10.1051/0004-6361/201321679 Astron. Astrophys. 562, A94.

[Connors et al. (2011)] Connors, M., Wiegert, P., Veillet, C., 2011. Earth's Trojan asteroid. http://dx.doi.org/10.1038/nature10233 Nature 475, 481–483.

[Dailey et al. (2010)] Dailey, J. et al., 2010. The WISE Moving Object Pipeline Subsystem: Design and implementation. In: http://adsabs.harvard.edu/abs/2010AAS...21640904D Bull. Am. Astron. Soc. 41, 817.

[De Val-Borro et al. (2012)] De Val-Borro, M. et al., 2012. An upper limit for the water outgassing rate of main-belt comet 176P/LINEAR observed with Herschel/HIFI. http://dx.doi.org/10.1051/0004-6361/201220169 Astron. Astrophys. 546, L4.

[Deb & Singh (2009)] Deb, S., & Singh, H.P., 2009. Light curve analysis of variable stars using Fourier decomposition and principal component analysis. http://dx.doi.org/10.1051/0004-6361/200912851 Astron. Astrophys. 507, 1729–1737.

[DeMeo et al. (2013)] DeMeo, F.E., Binzel, R.P., Carry, B., Polishook, D., Moskovitz, N.A., 2013. Unexpected D-type interlopers in the inner main-belt. http://arxiv.org/pdf/1312.2962.pdf arXiv:1312.2962v1.

[Denneau et al. (2013)] Denneau, L., Jedicke, R., Grav, T., Granvik, M., Kubica, J., Milani, A., Vereš, P., Wainscoat, R., Chang, D., Pierfederici, F., Kaiser, N., Chambers, K.C., Heasley, J.N., Magnier, E.A., Price, P.A., Myers, J., Kleyna, J., Hsieh, H., Farnocchia, D., Waters, C., Sweeney, W.H., Green, D., Bolin, B., Burgett, W.S., Morgan, J.S., Tonry, J.L., Hodapp, K.W., Chastel, S., Chesley, S., Fitzsimmons, A., Holman, M., Spahr, T., Tholen, D., Williams, G.V., Abe, S., Armstrong, J.D., Bressi, T.H., Holmes, R., Lister, T., McMillan, R.S., Micheli,

M., Ryan, E.V., Ryan, W.H., Scotti, J.V., 2013. The Pan-STARRS Moving Object Processing System. http://dx.doi.org/10.1086/670337 Publ. Astron. Soc. Pac. 125, 357–395.

[Drake et al. (2009)] Drake, A.J. et al., 2009. First results from the Catalina Real-Time Transient Survey. http://dx.doi.org/10.1088/0004-637X/696/1/870 Astrophys. J. 696,870–884.

[Elst et al. (1996)] Elst, E.W. et al., 1996. Comet P/1996 N2 (Elst-Pizarro). http://www.cbat.eps.harvard.edu/iauc/06400/06456.html IAU Circ. 6456.

[Elvis (2014)] Elvis, M., 2014. How many ore-bearing asteroids? http://dx.doi.org/10.1016/j.pss.2013.11.008 Plant. Space Sci. 91, 20–26.

[Elvis et al. (2011)] Elvis, M., McDowell, J., Hoffman, J.A., Binzel, R.P., 2011. Ultra-low delta-v objects and the human exploration of asteroids. http://dx.doi.org/10.1016/j.pss.2011.05.006 Plant. Space Sci. 59, 1408–1412.

[Emery et al. (2011)] Emery, J.P., Burr, D.M., Cruikshank, D.P, 2011. Near-infrared spectroscopy of Trojan asteroids: Evidence for two compositional groups. http://dx.doi.org/10.1088/0004-6256/141/1/25 Astron. J. 141, 25.

[Fornasier et al. (2004)] Fornasier, S. et al., 2004. Visible spectroscopic and photometric survey of L5 Trojans: investigation of dynamical families. http://dx.doi.org/10.1016/j.icarus.2004.06.015 Icarus 172, 221–232.

[Fornasier et al. (2007)] Fornasier, S. et al., 2007. Visible spectroscopic and photometric survey of Jupiter Trojans: Final results on dynamical families. http://dx.doi.org/10.1016/j.icarus.2007.03.033 Icarus 190, 622–642.

[French (1987)] French, L.M., 1987. Rotation properties of four L5 Trojans from CCD photometry. http://dx.doi.org/10.1016/0019-1035(87)90178-3 Icarus 72, 325–341.

[French et al. (2011)] French, L.M., Stephens, R.D., Lederer, S.M., Coley, D.R., Rohl, 2011. Preliminary results from a study of Trojan asteroids. http://adsabs.harvard.edu/abs/2011MPBu...38..116F Minor Planet Bulletin 38, 116–120.

[French et al. (2012)] French, L.M., Stephens, R.D., Coley, D.R., Megna, R., Wasserman, L., 2012. Photometry of 17 Jovian Trojan asteroids. http://adsabs.harvard.edu/abs/2012MPBu...39..183F Minor Planet Bulletin 39, 183–187.

[French et al. (2015)] French, L.M., Stephens, R.D., Coley, D.R., Wasserman, L., Sieben, J., 2015. Rotation lightcurves of small Jovian Trojan asteroids. http://dx.doi.org/10.1016/j.icarus.2015.03.026 Icarus 254, 1–17.

[Garradd et al. (2008)] Garradd, G.J., Sostero, G., Camilleri, P., Guido, E., Jacques, C., Pimentel, E., 2008. Comet C/2008 R1 (Garradd). http://www.cbat.eps.harvard.edu/iauc/08900/08969.html IAU Circ. 8969.

[Gehrels (1956)] Gehrels, T., 1956. Photometric studies of asteroids. V. The light-curve and phase function of 20 Massalia. http://dx.doi.org/10.1086/146166 Astrophys. J. 123, 331.

[Gehrels and Binzel (1984)] Gehrels, T., Binzel, R.P., 1984. The Spacewatch Camera. http://adsabs.harvard.edu/abs/1984MPBu...11....1G The Minor Planet Bulletin 11, 1–2.

[Geurts et al. (2006)] Geurts, P., Ernst., D., Wehenkel, L., 2006. Extremely randomized trees. http://dx.doi.org/10.1007/s10994-006-6226-1 Machine Learning 63, 3–42.

[Gibbs et al. (2012c)] Gibbs, A.R. et al., 2012. Comet P/2012 F5 (Gibbs). http://www.cbat.eps.harvard.edu/cbet/003000/CBET003069.txt Central Bureau Electronic Telegrams 3069.

[Gilbert and Wiegert (2010)] Gilbert, A.M., Wiegert, P.A., 2010. Updated results of a search for main-belt comets using the Canada-France-Hawaii Telescope legacy survey. http://dx.doi.org/10.1016/j.icarus.2010.07.016 Icarus 210, 998–999.

[Giorgini et al. (1996)] Giorgini, J.D. et al., 1996. JPL's on-line solar system data service. In: http://adsabs.harvard.edu/abs/1996DPS....28.2504G Bull. Am. Astron. Soc. 28, 1158.

[Goldreich et al. (2002))] Goldreich, P., Lithwick, Y., Sari, R., 2002. Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. http://dx.doi.org/10.1038/nature01227 Nature 420, 643–646.

[Gomes et al. (2005))] Gomes, R., Levison, H.F., Tsiganis, K., Morbidelli, A., 2005. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. http://dx.doi.org/10.1038/nature03676 Nature 435, 466–469.

[Granvik et al. (2012)] Granvik, M., Vaubaillon, J., Jedicke, R., 2012. The population of natural Earth satellites. http://dx.doi.org/10.1016/j.icarus.2011.12.003 Icarus 218, 262–277.

[Grav et al. (2011))] Grav, T. et al., 2011. WISE/NEOWISE observations of the Jovian Trojans: Preliminary results. http://dx.doi.org/10.1088/0004-637X/742/1/40 Astrophys. J. 742, 40.

[Grav et al. (2012))] Grav, T. et al., 2012. WISE/NEOWISE observations of the Jovian Trojan population: Taxonomy. http://dx.doi.org/10.1088/0004-637X/759/1/49 Astrophys. J. 759, 49.

[Green (2006)] Green, D.W., 2006. Comet 176P/LINEAR = (118401) LINEAR. http://adsabs.harvard.edu/abs/2010ASPC..434...28G IAU Circ. 8722.

[Grillmair et al. (2010)] Grillmair, C.J. et al., 2010. An overview of the Palomar Transient Factory pipeline and archive at the Infrared Processing and Analysis Center. In: http://adsabs.harvard.edu/abs/2010ASPC..434...28G Astron. Soc. Pacific Conf. Ser. 434, 28.

[Høg et al., 2000] Høg, E. et al., 2000. The Tycho-2 catalogue of the 2.5 million brightest stars. http://adsabs.harvard.edu//abs/2000A%26A...355L..27H Astron. Astrophys. 355, L27–L30.

[Hanuš & Ďurech (2012))] Hanuš, J., Ďurech, J., 2012. The potential of sparse photometric data in asteroid shape modeling. http://dx.doi.org/10.1016/j.pss.2012.01.014 Planet. Space Sci. 73, 75–79.

[Hanuš et al. (2013))] Hanuš, J. et al., 2013. Asteroids' physical models from combined dense and sparse photometry and scaling of the YORP effect by the observed obliquity distribution. http://dx.doi.org/10.1051/0004-6361/201220701 Astron. Astrophys. 551, A67.

[Hapke (2012))] Hapke, B., 2012. Theory of Reflectance and Emittance Spectroscopy, second ed. Cambridge University Press, Cambridge, UK.

[Harris (2008)] Harris, A.W., 2008. What Spaceguard did. http://dx.doi.org/10.1038/4531178a Nature 453, 1178–1179.

[Harris (2013)] Harris, A.W., 2013. Population of NEAs and survey completion. IAA Planetary Defense Conference. Flagstaff, AZ.

[Harris & Burns (1979)] Harris, A.W., Burns, J.A., 1979. Asteroid rotation: I. Tabulation and analysis of rates, pole positions and shapes. http://dx.doi.org/10.1016/0019-1035(79)90058-7 Icarus 40, 115–144.

[Harris & Young (1988)] Harris, A.W., Young, J.W., 1988. Observations of asteroid phase relations. In: http://adsabs.harvard.edu/abs/1988BAAS...20..865H Bull. Am. Astron. Soc. 20, 865.

[Harris & Young (1989)] Harris, A.W., Young, J.W., 1989. Asteroid lightcurve observations from 1979–1981. http://dx.doi.org/10.1016/0019-1035(89)90056-0 Icarus 81, 314–364.

[Harris et al. (1989a)] Harris, A.W. et al., 1989a. Photoelectric observations of asteroids 3, 24, 60, 261, and 863. http://dx.doi.org/10.1016/0019-1035(89)90015-8 Icarus 77, 171–186.

[Harris et al. (1989b)] Harris, A.W. et al., 1989b. Phase relations of high albedo asteroids - The unusual opposition brightening of 44 Nysa and 64 Angelina. http://dx.doi.org/10.1016/0019-1035(89)90057-2 Icarus 81, 365–374.

[Harris et al. (1992)] Harris, A.W., Young, J.W., Dockweiler, T., Gibson, J., 1992. Asteroid lightcurve observations from 1981. http://dx.doi.org/10.1016/0019-1035(92)90195-D Icarus 95, 115–147.

[Harris et al. (2012)] Harris, A.W., Warner, B.D., Pravec, P., Eds., 2012. Asteroid Lightcurve Derived Data V13.0. http://sbn.psi.edu/pds/resource/lc.html NASA Planetary Data System, EAR-A-5-DDR-DERIVED-LIGHTCURVE-V13.0..

[Harris et al. (2014)] Harris, A.W. et al., 2014. On the maximum amplitude of harmonics of an asteroid lightcurve. http://dx.doi.org/10.1016/j.icarus.2014.03.004 Icarus 235, 55–59.

[Helfenstein & Veverka (1989)] Helfenstein, P., Veverka, J., 1989. Physical characterization of asteroid surfaces from photometric analysis. In: http://adsabs.harvard.edu/abs/1989aste.conf..557H Binzel, R.P., Gehrels, T., Shapley Matthews, M.. (Eds.), Asteroids II. Univ. of Arizona Press, Tucson, pp. 557–593.

[Helin and Shoemaker (1979)] Helin, E.F., Shoemaker, E.M., 1979. The Palomar planet-crossing asteroid survey, 1973–1978. http://dx.doi.org/10.1016/0019-1035(79)90021-6 Icarus 40, 321–328.

[Hogg et al. (2010))] Hogg, D.W., Bovy, J., Lang, D., 2010. Data analysis recipes: Fitting a model to data. http://adsabs.harvard.edu/abs/2010arXiv1008.4686H arXiv:1008.4686.

[Hsieh (2009)] Hsieh, H.H., 2009. The Hawaii trails project: Comet-hunting in the main asteroid belt. http://dx.doi.org/10.1051/0004-6361/200912342 Astron. Astrophys. 505, 1297–1310.

[Hsieh and Jewitt (2006)] Hsieh, H.H., Jewitt, D.C., 2006. A population of comets in the main asteroid belt. http://dx.doi.org/10.1126/science.1125150 Science 312, 561–563.

[Hsieh et al. (2004)] Hsieh, H.H., Jewitt, D.C., Fernández, Y.R., 2004. The strange case of 133P/Elst-Pizarro: A comet among the asteroids. http://dx.doi.org/10.1086/383208 Astron. J. 127, 2997–3017.

[Hsieh et al. (2006b)] Hsieh, H.H., Jewitt, D.C., 2006b. Comet P/1999 RE70 = 118401. http://www.cbat.eps.harvard.edu/iauc/08700/08704.html IAU Circ. 8704.

[Hsieh et al. (2009a)] Hsieh, H.H., Jewitt, D.C., Fernandez, Y.R., 2009a. Albedos of main-belt comets 133P/Elst-Pizarro and 176P/LINEAR. http://dx.doi.org/10.1088/0004-637X/694/2/L111 Astrophys. J. 694, L111–L114.

[Hsieh et al. (2009b)] Hsieh, H.H., Jewitt, D.C., Ishiguro, M., 2009b. Physical properties of main-belt comet P/2005 U1 (Read). http://dx.doi.org/10.1088/0004-6256/137/1/157 Astron. J. 137, 157–168.

[Hsieh et al. (2010)] Hsieh, H.H., Jewitt, D.C., Lacerda, P., Lowry, S., Snodgrass, C., 2010. The return of activity in main-belt comet 133P/Elst-Pizarro. http://dx.doi.org/10.1111/j.1365-2966.2009.16120.x Mon. Not. R. Astron. Soc. 403, 363–377.

[Hsieh et al. (2011a)] Hsieh, H.H., Ishiguro, M., Lacerda, P., Jewitt, D.C., 2011a. Physical properties of main-belt comet 176P/LINEAR. http://dx.doi.org/10.1088/0004-6256/142/1/29 Astron. J. 142, 29.

[Hsieh et al. (2011b)] Hsieh, H.H., Meech, K.J., Pittichova, J., 2011b. Main-belt comet 238P/Read revisited. http://dx.doi.org/10.1088/2041-8205/736/1/L18 Astrophys. J. 736, L18.

[Hsieh et al. (2012a)] Hsieh, H.H. et al., 2012a. Observational and dynamical characterization of main-belt comet P/2010 R2 (La Sagra). http://dx.doi.org/10.1088/0004-6256/143/5/104 Astron. J. 143, 104.

[Hsieh et al. (2012b)] Hsieh, H.H. et al., 2012b. Discovery of main-belt comet P/2006 VW_{139} by Pan-STARRS 1. http://dx.doi.org/10.1088/2041-8205/748/1/L15 Astrophys. J. 748, L15.

[Hsieh et al. (2012c)] Hsieh, H.H., Denneau, L., Wainscoat, R., 2012c. Comet P/2012 T1 (PANSTARRS). http://www.cbat.eps.harvard.edu/cbet/003200/CBET003252.txt Central Bureau Electronic Telegrams 3252.

[Ivezić et al. (2001)] Ivezić, Ž. et al., 2001. Solar system objects observed in the Sloan Digital sky-survey commissioning data. http://dx.doi.org/10.1086/323452 Astron. J. 122, 2749–2784.

[Ivezić et al. (2002)] Ivezić, Ž. et al., 2002. Color confirmation of asteroid families. http://dx.doi.org/10.1086/344077 Astron. J. 124, 2943–2948.

[Ivezić et al. (2010)] Ivezić, Z., Juric, M., Lupton, R.H., Tabachnik, S., Quinn, T., and the SDSS Collaboration, 2010. SDSS Moving Object Catalog V3.0. EAR-A-I0035-3-SDSSMOC-V3.0. http://sbn.psi.edu/pds/resource/sdssmoc.html NASA Planetary Data System.

[Jedicke (1995)] Jedicke, R., 1995. Automated CCD scanning for near-Earth asteroids. http://adsabs.harvard.edu/abs/1995IAUS..167..157J IAU Symposium No. 167, New Developments in Array Technology and Applications. A. G. D. Philip, K. A. Janes and A. R. Upgren (Eds.). Kluwer Academic Publishers, Dordrecht.

[Jedicke and Metcalfe (1998)] Jedicke, R., Metcalfe, T.S., 1998. The orbital and absolute magnitude distributions of main belt asteroids. http://dx.doi.org/10.1006/icar.2001.6788 Icarus 131, 245–260.

[Jedicke et al. (2002)] Jedicke, R., Larsen, J., Spahr, T., 2002. Observational selection effects in asteroid surveys. In: http://adsabs.harvard.edu/abs/2002aste.conf...71J Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 71–87.

[Jedicke et al. (2013)] Jedicke, R., Schunova, E., Vereš, P., Denneau, L., 2013. ATLAS and PS2 asteroid retrieval mission target detection capability study. Prepared for NASA HEOMD and the Jet Propulsion Laboratory, 24 April.

[Jewitt (2012)] Jewitt, D.C., 2012. The active asteroids. http://dx.doi.org/10.1088/0004-6256/143/3/66 Astron. J. 143, 66.

[Jewitt & Luu (1990)] Jewitt, D.C., Luu, J., 1990. CCD spectra of asteroids II: The Trojans as spectral analogs of cometary nuclei. http://dx.doi.org/10.1086/115572 Astron. J. 100, 933–944.

[Jewitt and Guilbert-Lepoutre (2012)] Jewitt, D.C., Guilbert-Lepoutre, A., 2012. Limits to ice on asteroids 24 Themis and 65 Cybele. http://dx.doi.org/10.1088/0004-6256/143/1/21 Astron. J. 143, 21.

[Jewitt et al. (2009)] Jewitt, D.C., Yang, B., Haghighipour, N., 2009. Main-belt comet P/2008 R1 (Garradd). http://dx.doi.org/10.1088/0004-6256/137/5/4313 Astron. J. 137, 4313–4321.

[Jewitt et al. (2010)] Jewitt, D.C., Weaver, H., Agarwal, J., Mutchler, M., Drahus, M., 2010. A recent disruption of the main-belt asteroid P/2010 A2. http://dx.doi.org/10.1038/nature09456 Nature 467, 817–819.

[Jewitt et al. (2011)] Jewitt, D.C., Weaver, H., Mutchler, M., Larson, S., Agarwal, J., 2011. Hubble Space Telescope observations of main-belt comet (596) Scheila. http://dx.doi.org/10.1088/2041-8205/733/1/L4 Astrophys. J. 733, L4.

[Jewitt et al. (2012)] Jewitt, D.C., Weaver, H., Agarwal, J., Mutchler, M., Larson, S., 2012. HST measurements of main-belt comet 300163. In: http://adsabs.harvard.edu/abs/2012DPS....4430201J Bull. Am. Astron. Soc. 44, 302.01.

[Jones et al. (2006)] Jones, R. et al., 2006. The CFEPS Kuiper belt survey: Strategy and presurvey results. http://dx.doi.org/10.1016/j.icarus.2006.07.024 Icarus 186, 608–622.

[Jurić (2011)] Jurić, M., 2011. Large Survey Database: A distributed framework for storage and analysis of large data sets. In: http://adsabs.harvard.edu/abs/2011AAS...21743319J Bull. Am. Astron. Soc. 43, 433.19.

[Kaasalainen (2004)] Kaasalainen, M., 2004. Physical models of large number of asteroids from calibrated photometry sparse in time. http://dx.doi.org/10.1051/0004-6361:20048003 Astron. Astrophys. 422, L39–L42.

[Kaasalainen et al. (2001)] Kaasalainen, M., Torppa, J., Muinonen, K., 2001. Optimization methods for asteroid lightcurve inversion II. The complete inverse problem. http://dx.doi.org/10.1006/icar.2001.6674 Icarus 153, 37–51.

[Kaiser et al. (2002)] Kaiser, N. et al., 2002. Pan-STARRS: A large synoptic survey telescope. http://dx.doi.org/10.1117/12.457365 Proc. SPIE 4826, 154–164.

[Klein et al. (2009)] Klein, C., Bloom, J., Starr, D., Brewer, J., 2009. PyMPChecker. http://www.cacr.caltech.edu/hotwired2/program/presentations/PyMPCheckerKlein.pdf Hot-wiring the Transient Universe 2, April 29th 2009.

[Kron (1980)] Kron, R.G., 1980. Photometry of a complete sample of faint galaxies. http://adsabs.harvard.edu/abs/1980ApJS...43..305K Astrophys. J. Suppl. Ser. 43, 305.

[Kubica et al. (2007)] Kubica, J. et al., 2007. Efficient intra- and inter-night linking of asteroid detections using kd-trees. http://dx.doi.org/10.1016/j.icarus.2007.01.008 Icarus 189, 151–168.

[Kulkarni (2013)] Kulkarni, S.R., 2013. The intermediate Palomar Transient Factory (iPTF) begins. http://adsabs.harvard.edu/abs/2013ATel.4807....1K The Astronomer's Telegram, 4807.

[Lagrange(1772)] Lagrange, J.-L., 1772. Essai sur le problème des trois corps. Prix de l'Académie Royale des Sciences de Paris, tome IX, 230–331.

[Laher et al. (2014)] Laher, R.R., Surace, J., Grillmair, C.J., Ofek, E.O., Levitan, D., Sesar, B., van Eyken, J., Law, N., Helou, G., Hamam, N., Masci, F., Mattingly, S., Jackson, E., Hacopeans, E., Mi, W., Groom, S., Teplitz, H., Desai, V., Hale, D., Smith, R., Walters, R., Quimby, R., Kasliwal, M., Horesh, A., Bellm, E., Barlow, T., Waszczak, A., Prince, T.A., Kulkarni, S.R., 2014. IPAC image processing and data archiving for the Palomar Transient Factory. http://dx.doi.org/10.1086/677351 Publ. Astron. Soc. Pac. 121, 1395–1408.

[Larson (2007)] Larson, S., 2007. Current NEO surveys. http://dx.doi.org/10.1017/S1743921307003390 Near Earth Objects, our Celestial Neighbors - Opportunity and Risk, Proceedings of IAU Symposium 236. G.B. Valsecchi D., Vokrouhlický, A. Milani (Eds.). Cambridge: Cambridge University Press.

[Larson (2010)] Larson, S.M., 2010. 596 Scheila. http://www.cbat.eps.harvard.edu/iauc/09100/09188.html IAU Circ. 9188.

[Law et al. (2009)] Law, N. et al., 2009. The Palomar Transient Factory: System overview, performance, and first results. http://dx.doi.org/10.1086/648598 Publ. Astron. Soc. Pac. 121, 1395–1408.

[Law et al. (2010)] Law, N. et al., 2010. The Palomar Transient Factory Survey Camera: 1st year performance and results. http://dx.doi.org/10.1117/12.857400 Proc. SPIE, 7735, 77353M.

[Lazzaro et al. (2004)] Lazarro, D. et al., 2004. S^3OS^2: The visible spectroscopic survey of 820 asteroids. http://dx.doi.org/10.1016/j.icarus.2004.06.006 Icarus 172, 179–220.

[Levison et al. (2009)] Levison, H. F., Bottke, W. F., Jr., Gounelle, M., Morbidelli, A., Nesvorný, D., Tsiganis, K., 2009. Contamination of the asteroid belt by primordial trans-Neptunian objects. http://dx.doi.org/10.1038/nature08094 Nature 460, 364–366.

[Levitan et al. (2011)] Levitan, D. et al, 2011. PTF1 J071912.13+485834.0: An outbursting AM CVn system discovered by a synoptic survey. http://dx.doi.org/10.1088/0004-637X/739/2/68 Astrophys. J. 739, 68.

[Li and Thakar (2008)] Li, N., Thakar, A.R., 2008. CasJobs: A workflow environment designed for large scientific catalogs. http://dx.doi.org/10.1109/WORKS.2008.4723957 Third Workshop on Workflows in Support of Large-Scale Science, 2008.

[Li et al. (2006)] Li, J.Y. et al., 2006. Photometric analysis of 1 Ceres and surface mapping from HST observations. http://dx.doi.org/10.1016/j.icarus.2005.12.012 Icarus 182, 143–160.

[Licandro et al. (2011)] Licandro, J. et al., 2011. Testing the comet nature of main-belt comets: The spectra of 133P/Elst-Pizarro and 176P/LINEAR. http://dx.doi.org/10.1051/0004-6361/201117018 Astron. Astrophys. 532, A65.

[Licandro et al. (2013)] Licandro, J., Moreno, F., de Leòn, J., Tozzi, G.P., Lara, L.M., Cabrera-Lavers, A., 2013. Exploring the nature of new main-belt comets with the 10.4 m GTC telescope: (300163) 2006 VW139. http://dx.doi.org/10.1051/0004-6361/201220080 Astron. Astrophys. 550, A17.

[Lowry et al. (2007)] Lowry, S.C., Fitzsimmons, A., Pravec, P., Vokrouhlický, D., Boehnhardt, H., Taylor, P.A., Margot, J., Galád, A., Irwin, M., Irwin, J., Kusnirák, P., 2007. Direct detection of the asteroidal YORP effect. http://dx.doi.org/10.1126/science.1139040 Science 316, 272.

[Lu and Love (2005)] Lu, E.T., Love, S.G., 2005. Gravitational tractor for towing asteroids. http://dx.doi.org/10.1038/438177a Nature 438, 177–178.

[Lupishko et al. (1981)] Lupishko, D.F., Tupieva, F.A., Velichko, F.P., Kiselev, N.N., Chernova, G.P., 1981. UBV photometry of the asteroids 19 Fortuna and 29 Amphitrite. http://adsabs.harvard.edu/abs/1981AVest..15...25L Astronomicheskii Vestnik 15, 25–31.

[Luu and Jewitt (1992)] Luu, J., Jewitt, D.C., 1992. High resolution surface brightness profiles of near-Earth asteroids. http://dx.doi.org/10.1016/0019-1035(92)90134-S Icarus 97, 276–287.

[MacLennan and Hsieh (2012)] MacLennan, E.M., Hsieh, H.H., 2012. The nucleus of main-belt comet 259P/Garradd. http://dx.doi.org/10.1088/2041-8205/758/1/L3 Astrophys. J. 758, L3.

[Mahabal et al. (2008)] Mahabal, A., Djorgovski, S.G., Turmon, M., et al., 2008. Automated probabilistic classification of transients and variables. http://dx.doi.org/10.1002/asna.200710943 Astronomische Nachrichten 329, 288–291.

[Mainzer et al. (2011)] Mainzer, A., Grav, T., Bauer, J. et al. 2011. NEOWISE observations of near-Earth objects: Preliminary results. http://dx.doi.org/10.1088/0004-637X/743/2/156 Astrophys. J. 743:156.

[Mainzer et al. (2012)] Mainzer, A. et al, 2012. Physical parameters of asteroids estimated from the WISE 3-band data and NEOWISE post-cryogenic survey. http://dx.doi.org/10.1088/2041-8205/760/1/L12 Astrophys. J. 760, L12.

[Mann et al. (2007)] Mann, R.K., Jewitt, D.C., Lacerda, P., 2007. Fraction of contact binary Trojan asteroids. http://dx.doi.org/10.1086/520328 Astron. J. 134, 1133–1144.

[Marchis et al. (2006)] Marchis, F., et al., 2006. A low density of 0.8 g cm^{-3} for the Trojan binary asteroid 617 Patroclus. http://dx.doi.org/10.1038/nature04350 Nature 439 565–567.

[Marchis et al. (2014)] Marchis, F., et al., 2014. The puzzling mutual orbit of the binary Trojan asteroid 624 Hektor. http://dx.doi.org/10.1088/2041-8205/783/2/L37 Astrophys. J. Lett. 783, L37.

[Marsden (1986)] Marsden, B.G., 1986. Notes from the IAU General Assembly. http://www.minorplanetcenter.net/iau/ECS/MPCArchive/1985/MPC_19851227.pdf Minor Planet Circ. Nos. 10193 and 10194.

[Marsden (1996)] Marsden, B.G., 1996. Comet P/1996 N2 (Elst-Pizarro). http://www.cbat.eps.harvard.edu/iauc/06400/06457.html IAU Circ. 6457.

[Martin et al. (2005))] Martin, D.C. et al., 2005. The *Galaxy Evolution Explorer*: A space ultraviolet survey mission. http://dx.doi.org/10.1086/426387 Astrophys. J. 619, 1.

[Masci (2013)] Masci, F., 2013. PTFIDE: Palomar Transient Factory Image Differencing and Extraction - Algorithms and module usage. Document available at http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/ptfide-v4.0.pdf http://web.ipac.caltech.edu/staff/fmasci/home/miscscience/ptfide-v4.0.pdf.

[Masci et al. (2014)] Masci, F.J., Hoffman, D.I., Grillmair, C.J., Cutri, R.M., 2014. Automated classification of periodic variable stars detected by the Wide-field In-

frared Survey Explorer. http://dx.doi.org/10.1088/0004-6256/148/1/21 Astron. J. 148, 21.

[Masiero et al. (2009)] Masiero, J., Jedicke, R., Ďurech, J., Gwym, S., Denneau, L, Larsen, J., 2009. The Thousand Asteroid Light Curve Survey. http://dx.doi.org/10.1016/j.icarus.2009.06.012 Icarus 204, 145–171.

[Masiero et al. (2011)] Masiero, J. et al., 2011. Main-belt asteroids with WISE/NEOWISE. I. Preliminary albedos and diameters. http://dx.doi.org/10.1088/0004-637X/741/2/68 Astrophys. J. 741, 68.

[Masiero et al. (2012)] Masiero, J. et al., 2012. Preliminary analysis of WISE/NEOWISE 3-band cryogenic and post-cryogenic observations of main-belt asteroids. http://dx.doi.org/10.1088/2041-8205/759/1/L8 Astrophys. J. 759, L8.

[Masiero et al. (2014)] Masiero, J. et al., 2014. Main-belt asteroids with WISE/NEOWISE: Near-infrared albedos. http://dx.doi.org/10.1088/0004-637X/791/2/121 Astrophys. J. 791, 121.

[Massey (1951))] Massey, F.J., 1951. The Kolmogorov-Smirnov test for goodness of fit. http://dx.doi.org/10.1080/01621459.1951.10500769 Journal of the American Statistical Association 46, 68–78.

[Matson et al. (1986))] Matson, D.L., Veeder, G.J., Tedesco, E.F., Lebofsky, L.A., Walker, R.G., 1986. The IRAS survey of asteroids. http://dx.doi.org/10.1016/0273-1177(86)90210-3 Advances in Space Research 6, 47–56.

[McMillan (2000)] McMillan, R.S., Perry, M.L., Bressi, T.H., Montani, J.L., Tubbiolo, A.F., Read, M.T., 2000. Progress on the Spacewatch 1.8-m telescope and upgrade of the Spacewatch 0.9-m telescope. In: http://adsabs.harvard.edu/abs/2000DPS....32.2612M Bull. Am. Astron. Soc. 32, 1042.

[Melita et al. (2010)] Melita, M.D., Duffard, R., Williams, I.P., Jones, D.C., Licandro, J., Ortiz, J.L., 2010. Lightcurves of 6 Jupiter Trojan asteroids. http://dx.doi.org/10.1016/j.pss.2010.03.009 Planet. Space Sci. 58, 1035–1039.

[Miller et al. (2014)] Miller, A., Bloom, J.S., Richards, J., et al. 2014. A machine learning method to infer fundamental stellar parameters from photometric light curves. http://dx.doi.org/10.1088/0004-637X/798/2/122 Astrophys. J. 798, 122.

[Morbidelli et al. (2005))] Morbidelli, A., Levison, H.F., Tsiganis, K., Gomes, R., 2005. Chaotic capture of Jupiter's Trojan asteroids in the early solar system. http://dx.doi.org/10.1038/nature03540 Nature 435, 462–465.

[Morbidelli et al. (2012)] Morbidelli, A., Lunine, J.I., O'Brien, D.P., Raymond, S.N., Walsh, K.J., 2012. Building terrestrial planets. http://dx.doi.org/10.1146/annurev-earth-042711-105319 Annu. Rev. Earth Planet. Sci. 40, 251–275.

[Moreno et al. (2011)] Moreno, F. et al., 2011. The dust environment of main-belt comet P/2010 R2 (La Sagra). http://dx.doi.org/10.1088/2041-8205/738/1/L16 Astrophys. J. 738, L16.

[Moreno et al. (2012)] Moreno, F., Licandro, J., Cabrera-Lavers, A., 2012. A short-duration event as the cause of dust ejection from main-belt comet P/2012 F5 (Gibbs). http://dx.doi.org/10.1088/2041-8205/761/1/L12 Astrophys. J. 761, L12.

[Morissey et al. (2005))] Morissey, P. et al., 2005. The on-orbit performance of the *Galaxy Evolution Explorer*. http://dx.doi.org/10.1086/424734 Astrophys. J. 619, 1.

[Morissey et al. (2007))] Morissey, P. et al., 2007. The calibration and data products of *GALEX*. http://dx.doi.org/10.1086/520512 Astrophys. J. Suppl. Ser. 173, 682–697.

[Mottola et al. (2011))] Mottola, S. et al., 2011. Rotational properties of Jupiter Trojans. I. Light curves of 80 objects. http://dx.doi.org/10.1088/0004-6256/141/5/170 Astron. J. 141, 170.

[Mottola et al. (2014))] Mottola, S. et al., 2014. The spin rate distribution of Jupiter Trojans. http://adsabs.harvard.edu/abs/2014MSAIS..26...47M Memorie della Societa Astronomica Italiana Supplement 26, 47.

[Mueller et al. (2011)] Mueller, M., Delbo, M., Hora, J.L., Trilling, D.E., Bhattacharya, B., Bottke, W.F., Chesley, S., Emery, J.P., Fazio, G., Harris, A.W., Mainzer, A., Mommert, M., Penprase, B., Smith, H.A., Spahr, T.B., Stansberry, J.A., Thomas, C.A., 2011. ExploreNEOs III - physical characterization of 65 potential spacecraft target asteroids. http://dx.doi.org/10.1088/0004-6256/141/4/109 Astron. J. 141, 109.

[Muinonen et al. (2010))] Muinonen, K., Belskaya, I.N., Cellino, A., Delbò, M., Levasseur-Regourd, A., Penttil a, A., Tedesco, E.F., 2010. A three-parameter magnitude phase function for asteroids. http://dx.doi.org/10.1016/j.icarus.2010.04.003 Icarus 209, 542–555.

[Muller et al. (2010))] Muller, M. et al., 2010. Eclipsing binary Trojan asteroid Patroclus: Thermal inertia from Spitzer observations. http://dx.doi.org/10.1016/j.icarus.2009.07.043 Icarus 205, 505–515.

[Murray and Dermott (1999)] Murray, C.D., Dermott, S.F., 1999. Solar system dynamics. Cambridge: Cambridge Univ. Press. pg. 71.

[Nesvorný & Morbidelli (2012))] Nesvorný, D., Morbidelli, A., 2012. Statistical study of the early solar system's instability with four, five, and six giant planets. http://dx.doi.org/10.1088/0004-6256/144/4/117 Astron. J. 144, 117.

[Nesvorný et al. (2013))] Nesvorný, D., Vokrouhlický, D., Morbidelli, A., 2013. Capture of Trojans by jumping Jupiter. http://dx.doi.org/10.1088/0004-637X/768/1/45 Astrophys. J. 768, 45.

[Nomen et al. (2011)] Nomen, J. et al., 2010. Comet P/2010 R2 (La Sagra). http://www.cbat.eps.harvard.edu/iauc/09100/09169.html IAU Circ. 9169.

[Novaković et al. (2012)] Novaković, B., Hsieh, H.H., Cellino, A., 2012. P/2006 VW_{139}: A main-belt comet born in an asteroidal collision? http://dx.doi.org/10.1111/j.1365-2966.2012.21329.x Mon. Not. R. Astron. Soc. 424, 1432–1441.

[Obama (2010)] Obama, B., 2010. Speech at Kennedy Space Center, 15 April.

[Ofek (2012)] Ofek, E.O., 2012. Sloan Digital sky-survey observations of Kuiper belt objects: Colors and variability. http://dx.doi.org/10.1088/0004-637X/749/1/10 Astrophys. J. 749, 10.

[Ofek et al. (2011)] Ofek, E.O. et al., 2011. A Very Large Array search for 5 GHz radio transients and variables at low galactic latitudes. http://dx.doi.org/10.1088/0004-637X/740/2/65 Astrophys. J. 740, 65.

[Ofek et al. (2012a)] Ofek, E.O. et al., 2012a. The Palomar Transient Factory photometric calibration. http://dx.doi.org/10.1086/664065 Publ. Astron. Soc. Pac. 124, 62–73.

[Ofek et al. (2012b)] Ofek, E.O. et al., 2012b. The Palomar Transient Factory photometric catalog 1.0. http://dx.doi.org/10.1086/666978 Publ. Astron. Soc. Pac. 124, 854–860.

[Ofek et al. (2014)] Ofek, E.O., 2014. MATLAB package for astronomy and astrophysics. http://adsabs.harvard.edu/abs/2014ascl.soft07005O Astrophysics Source Code Library, record ascl:1407.005.

[Oszkiewicz et al. (2011)] Oszkiewicz, D.A. et al., 2011. Online multi-parameter phase-curve fitting and application to a large corpus of asteroid photometric data. http://dx.doi.org/10.1016/j.jqsrt.2011.03.003 J. Quant. Spectros. Radiat. Trans. 112, 1919–1929.

[Oszkiewicz et al. (2012)] Oszkiewicz, D.A. et al., 2012. Asteroid taxonomic signatures from photometric phase curves. http://dx.doi.org/10.1016/j.icarus.2012.02.028 Icarus 219, 283–296.

[Owen (2008)] Owen, T., 2008. The contributions of comets to planets, atmospheres, and life: Insights from Cassini-Huygens, Galileo, Giotto, and inner planet missions. http://dx.doi.org/10.1007/s11214-008-9306-7 Space Sci. Rev. 138, 301–316.

[Park et al. (2013)] Park, M., Oh, H.-S., Kim, D., 2013. Classification of variable stars using thick-pen transform method. http://dx.doi.org/10.1086/670671 Publ. Astron. Soc. Pac. 125, 470–476.

[Parker et al. (2002)] Parker, J.W., Stern, A., Thomas, P.C., Festou, M.C., Merline, W.J., Young, E.F., Binzel, R.P., Lebofsky, L.A., 2002. Analysis of the first disk-resolved images of Ceres from ultraviolet observations with the Hubble Space Telescope. http://dx.doi.org/10.1086/338093 Astron. J. 123, 549–557.

[Parker et al. (2008)] Parker, A., Ivezić, Ž., Jurić, M., Lupton, R., Sekora, M.D., Kowalski, A., 2008. The size distributions of asteroid families in the SDSS Moving Object Catalog 4. http://dx.doi.org/10.1016/j.icarus.2008.07.002 Icarus 198, 138–155.

[Pickles (1998)] Pickles, A.J., 1998. A stellar flux library: 1150–2500 Å. http://dx.doi.org/10.1086/316197 Publ. Astron. Soc. Pac. 110, 749.

[Polishook & Brosch (2009)] Polishook, D., Borsch, N., 2009. Photometry and spin rate distribution of small-sized main belt asteroids. http://dx.doi.org/10.1016/j.icarus.2008.10.020 Icarus 199, 319–332.

[Polishook et al. (2012)] Polishook, D. et al., 2012. Asteroid rotation periods from the Palomar Transient Factory survey. http://dx.doi.org/10.1111/j.1365-2966.2012.20462.x Mon. Not. R. Astron. Soc. 421, 2094–2108.

[Pravdo et al. (1999)] Pravdo, S.H., Rabinowitz, D., Helin, E.F., Lawrence, K.J., Bambery, R.J., Clark, C.C., Groom, S.L., Levin, S., Lorre, J., Shaklan, S.B., Kervin, P., Africano, J.A., Sydney, P., Soohoo, V., 1999. The Near-Earth Asteroid Tracking (NEAT) Program - An automated system for telescope control, wide-field imaging, and object detection. http://dx.doi.org/10.1086/300769 Astron. J. 117, 1616–1633.

[Pravec & Harris (2000)] Pravec, P., Harris, A.W., 2000. Fast and slow rotation of asteroids. http://dx.doi.org/10.1006/icar.2000.6482 Icarus 148, 12–20.

[Pravec et al. (2002)] Pravec, P., Harris, A.W., Michałowski, T., 2002. Asteroid rotations. In: http://adsabs.harvard.edu/abs/2002aste.conf..113P Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 113–122.

[Pravec et al. (2002)] Pravec, P., Harris, A.W., Michalowski, T., 2002. Asteroid rotations. In: http://adsabs.harvard.edu/abs/2002aste.conf..113P Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 113–122.

[Pravec et al. (2008)] Pravec, P., Harris, A.W., Vokrouhlický, D., et al., 2008. Spin rate distribution of small asteroids. http://adsabs.harvard.edu/abs/2008Icar..197..497P Icarus 197, 497–504.

[Prialnik and Rosenberg (2009)] Prialnik, D., Rosenberg, E.D., 2009. Can ice survive in main-belt comets? Long-term evolution models of comet 133P/Elst-Pizarro. http://dx.doi.org/10.1111/j.1745-3933.2009.00727.x Mon. Not. R. Astron. Soc. 399, L79–L83.

[Rabinowitz (1991)] Rabinowitz, D., 1991. Detection of Earth-approaching asteroids in near real time. http://dx.doi.org/10.1086/115785 Astron. J. 101, 1518–1529.

[Rabinowitz et al. (1993)] Rabinowitz, D., Gehrels, T., Scotti, J.V., McMillan, R.S., Perry, M.L., Wisniewski, W., larson, S.M., Howell, E.S., Mueller, B.E., 1993. Evi-

dence for a near-Earth asteroid belt. http://dx.doi.org/10.1038/363704a0 Nature 363, 704–706.

[Rabinowitz et al. (2000)] Rabinowitz, D., Helin, E., Lawrence, K., Pravdo, S., 2000. A reduced estimate of the number of kilometer-sized near-Earth asteroids. http://dx.doi.org/10.1038/35003128 Nature 403, 165–166.

[Rau et al. (2009)] Rau, A. et al., 2009. Exploring the optical transient sky with the Palomar Transient Factory. http://dx.doi.org/10.1086/605911 Publ. Astron. Soc. Pac. 103, 1334–1351.

[Read (2005)] Read, M.T., 2005. Comet P/2005 U1 (Read). http://www.cbat.eps.harvard.edu/iauc/08600/08624.html IAU Circ. 8624.

[Robert (2011)] Robert, F., 2011. A distinct source for lunar water? http://dx.doi.org/10.1038/ngeo1066 Nature Geoscience 4, 74–75.

[Roettger and Buratti (1994)] Roettger, E.E., Buratti, B.J., 1994. Ultraviolet spectra and geometric albedos of 45 asteroids. http://dx.doi.org/10.1006/icar.1994.1201 Icarus 112, 496–512.

[Roig et al. (2008)] Roig, F., Ribeiro, A.O., Gil-Hutton, R., 2008. Taxonomy of asteroid families among the Jupiter Trojans: Comparison between spectroscopic data and the Sloan Digital sky-survey colors. http://dx.doi.org/10.1051/0004-6361:20079177 Astron. Astrophys. 483, 911–931.

[Rousselot et al. (2011)] Rousselot, P., Dumas, C., Merlin, F., 2011. Near-infrared spectroscopy of 133P/Elst-Pizarro. http://dx.doi.org/10.1016/j.icarus.2010.08.025 Icarus 211, 553–558.

[Ruprecht et al. (2014)] Ruprecht, J.D., Stuart, J.S., Woods, D.F., Shah, R.Y., 2014. Detecting small asteroids with the Space Surveillance Telescope. http://dx.doi.org/10.1016/j.icarus.2014.06.012 Icarus 239, 253–259.

[Sagiv et al. (2014)] Sagiv, I., Gal-Yam, A., Ofek, E.O., Waxman, E., Aharonson, O., Kulkarni, S.R., Nakar, E., Maoz, D., Trakhtenbrot, B., Phinney, S., 2014. Science with a wide-field UV transient explorer. http://dx.doi.org/10.1088/0004-6256/147/4/79 Astron. J. 147, 79.

[Salo (1987)] Salo, H., 1987. Numerical simulations of collisions between rotating particles. http://dx.doi.org/10.1016/0019-1035(87)90073-X Icarus 70, 37–51.

[Sanchez et al. (2012)] Sanchez, J.A., Reddy, V., Nathues, A., Cloutis, E.A., Mann, P., Hiesinger, H., 2012. Phase reddening on near-Earth asteroids: Implications for mineralogical analysis, space weathering and taxonomic classification. http://dx.doi.org/10.1016/j.icarus.2012.04.008 Icarus 220, 36–50.

[Schaefer et al. (2010)] Schaefer, M.W., Schaefer, B.E., Rabinowitz, D.L., Toutellotte, S.W., 2010. Phase curves of nine Trojan asteroids over a wide range of phase angles. http://dx.doi.org/10.1016/j.icarus.2009.11.031 Icarus 207, 699–713.

[Schorghofer (2008)] Schorghofer, N., 2008. The lifetime of ice on main belt asteroids. http://dx.doi.org/10.1086/588633 Astrophys. J. 682, 697–705.

[Scotti et al. (1991)] Scotti, J.V., Rabinowitz, D., Marsden, B.G., 1991. Near miss of the Earth by a small asteroid. http://dx.doi.org/10.1038/354287a0 Nature 354, 287–289.

[Seager (2010)] Seager, S., 2010. http://adsabs.harvard.edu/abs/2010eapp.book.....S Exoplanet Atmospheres: Physical Processes. Princeton, NJ: Princeton Univ. Press. pp. 46.

[Shevchenko (1997)] Shevchenko, V.G., 1997. Analysis of asteroid brightness-phase relations. http://adsabs.harvard.edu/abs/1997SoSyR..31..219S Solar System Research 31, 219–224.

[Shevchenko et al. (2012)] Shevchenko, V.G. et al., 2012. Opposition effect of Trojan asteroids. http://dx.doi.org/10.1016/j.icarus.2011.11.001 Icarus 217, 202–208.

[Shevchenko et al. (2013)] Shevchenko, V.G. et al., 2013. Influence on magnitude phase angle dependence of asteroid 596 Scheila. http://adsabs.harvard.edu/abs/2013LPI....44.2706S Lunar Planet. Sci. Conf. Contribution No. 1719, 2706.

[Shoemaker & Helin (1978)] Shoemaker, E.M., Helin, E.F., 1978. Earth-approaching asteroids as targets for exploration. In: D. Morrison and W. C. Wells (Eds.), http://adsabs.harvard.edu/abs/1978NASCP2053..245S NASA Conference Publication, Volume 2053 of NASA Conference Publication, pp. 245.

[Silber et al. (2009)] Silber, E.A., ReVelle, D.O., Brown, P.G., Edwards, W.N., 2009. An estimate of the terrestrial influx of large meteoroids from infrasonic measurements. http://dx.doi.org/10.1029/2009JE003334 J. Geophys. Res. 114, E08006.

[Singer et al. (2013)] Singer, L.P., Cenko, S.B., Kasliwal, M.M., Perley, D.A., Ofek, E.O., Brown, D.A., Nugent, P.E., Kulkarni, S.R., Corsi, A., Frail, D.A., Bellm, E., Mulchaey, J., Arcavi, I., Barlow, T., Bloom, J.S., Cao, Y., Gehrels, N., Horesh, A., Masci, F.J., McEnery, J., Rau, A., Surace, J., Yaron, O. 2013. Discovery and redshift of an optical afterglow in 71 deg^2: iPTF13bxl and GRB 130702A. http://dx.doi.org/10.1088/2041-8205/776/2/L34 Astrophys. J. Lett. 776:L34.

[Skrutskie et al. (2006)] Skrutskie, M.F., et al., 2006. The Two Micron All sky-survey (2MASS). http://dx.doi.org/10.1086/498708 Astron. J. 131, 1163–1183.

[Snodgrass et al. (2010)] Snodgrass, C. et al., 2010. A collision in 2009 as the origin of the debris trail of asteroid P/2010 A2. http://dx.doi.org/10.1038/nature09453 Nature 467, 814–816.

[Sonnett et al. (2011)] Sonnett, S., Kleyna, J., Jedicke, R., Masiero, J., 2011. Limits on the size and orbital distribution of main belt comets. http://dx.doi.org/10.1016/j.icarus.2011.08.001 Icarus 215, 534–546.

[Sonnett et al. (2015)] Sonnett, S. et al., 2015.Binary candidates in the Jovian Trojan and Hilda populations from NEOWISE light curves. http://dx.doi.org/10.1088/0004-637X/799/2/191 Astrophys. J. 799:191.

[Steinberg & Sari (2015)] Steinberg, E., Sari, R., 2015. Spins of large asteroids: A hint of a primordial distribution in their spin rates. http://dx.doi.org/10.1088/0004-6256/149/4/124 Astron. J. 149, 124.

[Stern et al. (2011)] Stern, S.A. et al. 2011. Ultraviolet discoveries at asteroid (21) Lutetia by the *Rosetta* Alice ultraviolet spectrograph. http://dx.doi.org/10.1088/0004-6256/141/6/199 Astron. J. 141:199.

[Stetson (1996)] Stetson, P.B., 1996. on the automatic determination of light-curve parameters for Cepheid variables. http://dx.doi.org/10.1086/133808 Publ. Astron. Soc. Pac. 108, 851.

[Stevenson et al. (2012)] Stevenson, R., Kramer, E.A., Bauer, J.M., Masiero, J.R., Mainzer, A.K., 2012. Characterization of active main belt object P/2012 F5 Gibbs: A possible impacted asteroid. http://dx.doi.org/10.1088/0004-637X/759/2/142 Astrophys. J. 759, 142.

[Stokes et al. (2000)] Stokes, G.H., Evans, J.B., Viggh, E.M., Shelly, F.C., Pearce, E.C., 2000. Lincoln Near-Earth Asteroid program (LINEAR). http://dx.doi.org/10.1006/icar.2000.6493 Icarus 148, 21–28.

[Stokes et al. (2002)] Stokes, G.H., Evans, J.B., larson, S.M., 2002. Near-Earth asteroid search programs. In: http://adsabs.harvard.edu/abs/2002aste.conf...45S Bottke, W.F., Jr., Cellino, A., Paolicchi, P., Binzel, R. (Eds.), Asteroids III. Univ. of Arizona Press, Tucson, pp. 45–54.

[Stuart (2001)] Stuart, J.S., 2001. A near-Earth asteroid population estimate from the LINEAR survey. http://dx.doi.org/10.1126/science.1065318 Science 294, 1691.

[Sykes et al. (2000)] Sykes, M.V. et al., 2000. The 2MASS Asteroid and Comet Survey. http://dx.doi.org/10.1006/icar.2000.6366 Icarus 146, 161–175.

[Sykes et al. (2010)] Sykes, M.V., Cutri, R,M., Skrutskie, M.F., Fowler, J.W., Tholen, D.J., Painter, P.E., Nelson, B., Kirkpatrick, D.J., 2010. 2MASS Asteroid and Comet Survey V2.0. EAR-A-I0054/I0055-5-2MASS-V2.0. http://sbn.psi.edu/pds/resource/2mass.html NASA Planetary Data System.

[Szabó et al. (2007)] Szabó, G.M., Ivezić, Ž., Jurić, M., Lupton, R., 2007. The properties of Jovian Trojan asteroids listed in the SDSS Moving Object Catalogue 3. http://dx.doi.org/10.1111/j.1365-2966.2007.11687.x Mon. Not. R. Astron. Soc. 377, 1393–1406.

[Tedesco (1995)] Tedesco, E.F., 1995. UBV Mean Asteroid Colors. EAR-A-5-DDR-UBV-MEAN-VALUES-V1.2. http://sbn.psi.edu/pds/resource/ubvmean.html NASA Planetary Data System.

[Tedesco et al. (2002a)] Tedesco, E.F., Noah, P.V., Noah, M., Price, S.D., 2002. The Supplemental IRAS Minor Planet Survey. http://dx.doi.org/10.1086/338320 Astron. J. 123, 1056–1085.

[Tedesco et al. (2002b)] Tedesco, E.F.,Egan, M.P., Price, S.D., 2002. The Midcourse Space Experiment Infrared Minor Planet Survey. http://dx.doi.org/10.1086/340960 Astron. J. 124, 583–591.

[Tholen (1989)] Tholen, D.J., 1989. Asteroid taxonomic classifications. In: Asteroids II (R. P. Binzel, T. Gehrels, and M. S. Matthews, Eds.), pp. 1139–1150. Univ. of Arizona Press, Tucson.

[Toth (2006)] Toth, I., 2006. The quasi-Hilda subgroup of ecliptic comets: An update. http://dx.doi.org/10.1051/0004-6361:20053492 Astron. Astrophys. 448, 1191–1196.

[Tsiganis et al. (2005)] Tsiganis, K., Gomes, R., Morbidelli, A., Levison, H.F., 2005. Origin of the orbital architecture of the giant planets of the solar system. http://dx.doi.org/10.1038/nature03539 Nature 435, 459–461.

[Usui et al. (2011)] Usui, F. et al., 2011. Asteroid catalog using AKARI: AKARI/IRC Mid-infrared asteroid survey. http://dx.doi.org/10.1093/pasj/63.5.1117 Publ. Astron. Soc. Japan 63, 1117–1138.

[Usui et al. (2014)] Usui, F. et al., 2014. A comparative study of infrared asteroid surveys: IRAS, AKARI, and WISE. http://dx.doi.org/10.1093/pasj/psu037 Publ. Astron. Soc. Japan 66, 1–11.

[Vereš et al. (2012)] Vereš, P., Jedicke, R., Denneau, L., Wainscoat, R., Holman, M., Lin, H.-W., 2012. Improved asteroid astrometry and photometry with trail-fitting. http://dx.doi.org/10.1086/668616 Publ. Astron. Soc. Pac. 124, 1197–1207.

[Wagner et al. (1987)] Wagner, J.K, Hapke, B.W., Wells, E.N., 1987. Atlas of reflectance spectra of terrestrial, lunar, and meteoritic powders and frosts from 92 tp 1800 nm. http://dx.doi.org/10.1016/0019-1035(87)90003-0 Icarus 69, 14–28.

[Walsh et al. (2011)] Walsh, K.J., Morbidelli, A., Raymond, S.N., O'Brien, D.P., Mandell, A.M., 2011. A low mass for Mars from Jupiter's early gas-driven migration. http://dx.doi.org/10.1038/nature10201 Nature 475, 206–209.

[Warner & Stevens (2011)] Warner, B.D., Stephens, R.D., 2011. Lightcurve analysis for a trio of asteroids. http://adsabs.harvard.edu/abs/2011MPBu...38..110W Minor Planet Bulletin 38, 110–111.

[Warner et al. (2009)] Warner, B., Harris, A.W., Pravec, P., 2009. The asteroid lightcurve database. http://dx.doi.org/10.1016/j.icarus.2009.02.003 Icarus 202, 134–146.

[Warner et al. (2009)] Warner, B., Harris, A.W., Pravec, P., 2009. The asteroid lightcurve database. http://dx.doi.org/10.1016/j.icarus.2009.02.003 Icarus 202, 134–146.

[Waszczak et al. (2011)] Waszczak, A., Ofek, E.O., Polishook, D., 2011. 2011 CR_{42}. http://www.cbat.eps.harvard.edu/cbet/002800/CBET002823.txt Central Bureau Electronic Telegrams 2823.

[Waszczak et al. (2013a)] Waszczak, A., Ofek, E.O., Scotti, J.V., Kowalski, R.A., Williams, G.V., 2013. Comet C/2012 LP_{26} (Palomar). http://www.cbat.eps.harvard.edu/cbet/003400/CBET003408.txt Central Bureau Electronic Telegrams 3408.

[Waszczak et al. (2013b)] Waszczak., A. et al., 2013. Main-belt comets in the Palomar Transient Factory survey: I. The search for extendedness. http://dx.doi.org/10.1093/mnras/stt951 Mon. Not. R. Astron. Soc. 433, 3115–3132.

[Waszczak et al. (2015)] Waszczak., A. et al., 2015. Asteroid lightcurves from the Palomar Transient Factory survey: Rotation rates and phase functions from sparse photometry. http://adsabs.harvard.edu/abs/2015arXiv150404041W Astron. J. in press, arXiv:1504.04041.

[Weaver et al. (2010)] Weaver, H.A., Feldman, P.D., Merline, W.J., et al., 2010. Ultraviolet and visible photometry of asteroid 21 Lutetia using the Hubble Space Telescope. http://dx.doi.org/10.1051/0004-6361/200913950 Astron. Astrophys. 518, A4.

[Williams (2012)] Williams., G.V., 2012. Minor planet astrophotometry. http://adsabs.harvard.edu/abs/2012PhDT.....7W Ph.D. Thesis, Open University, Milton Keynes, U.K..

[Woźniak et al. (2004)] Woźniak, P.R., Williams, S.J., Vestrand, W.T., Gupta, V., 2004. Identifying red variables in the Northern Sky Variability Survey. http://dx.doi.org/10.1086/425526 Astron. J. 128, 2965–2976.

[Wong et al. (2014)] Wong, I., Brown, M.E., Emery, J.P., 2014. The differing magnitude distributions of the two Jupiter Trojan color populations. http://dx.doi.org/10.1088/0004-6256/148/6/112 Astron. J. 148, 112.

[Wright et al. (2010)] Wright, E.L. et al., 2010. The Wide-field Infrared Survey Explorer (WISE): Mission description and initial on-orbit performance. http://dx.doi.org/10.1088/0004-6256/140/6/1868 Astron. J. 140, 1868–1881.

[Xu et al. (1995)] Xu, S.,Binzel, R.P., Burbine, T.H., Bus, S.J., 1995. Small Main-Belt Asteroid Spectroscopic Survey. http://dx.doi.org/10.1006/icar.1995.1075 Icarus 115, 1–35.

[Xu et al. (1996)] Xu, S.,Binzel, R.P., Burbine, T.H., Bus, S.J., 1996. Small Main-Belt Asteroid Survey. EAR-A-M3SPEC-3-RDR-SMASS-V2.1. http://sbn.psi.edu/pds/resource/smass1.html NASA Planetary Data System.

[Yang et al. (2002)] Yang, B., Zhu, J., Song, Y., 2002. Ghost images in Schmidt CCD photometry. http://dx.doi.org/10.1088/1009-9271/2/5/474 Chin. J. Astron. Astrophys. 2, 474–480.

[York et al. (2000)] York, D.G. et al, 2000. The Sloan Digital sky-survey: Technical summary. http://dx.doi.org/10.1086/301513 Astron. J. 120, 1579–1587.

[Zacharias (2010)] Zacharias, N., 2010. UCAC3 pixel processing. http://dx.doi.org/10.1088/0004-6256/139/6/2208 Astron. J. 139, 2208–2217.

[Zellner et al. (1985)] Zellner, B., Tholen, D.J., Tedesco, E.F., 1998. The Eight-Color Asteroid Survey - Results for 589 minor planets. http://dx.doi.org/10.1016/0019-1035(85)90133-2 Icarus 61, 355–416.

[Zellner et al. (2009)] Zellner, B., Tholen, D.J., Tedesco, E.F., 20098. Eight Color Asteroid Survey. EAR-A-2CP-3-RDR-ECAS-V4.0. http://sbn.psi.edu/pds/resource/ecas.html NASA Planetary Data System.